Syracuse University

SURFACE

Dissertations - ALL

SURFACE

5-14-2017

DIRECTED EVOLUTION FOR THE DESIGN OF NEW CATALYSTS

Tiffany Dunston
Syracuse University

Follow this and additional works at: https://surface.syr.edu/etd

Part of the Physical Sciences and Mathematics Commons

Recommended Citation

Dunston, Tiffany, "DIRECTED EVOLUTION FOR THE DESIGN OF NEW CATALYSTS" (2017). *Dissertations - ALL*. 708.
https://surface.syr.edu/etd/708

This Dissertation is brought to you for free and open access by the SURFACE at SURFACE. It has been accepted for inclusion in Dissertations - ALL by an authorized administrator of SURFACE. For more information, please contact surface@syr.edu.

ABSTRACT

Introducing new catalytic function into existing proteins provide ways of understanding the fundamental principles of enzyme catalysis and has gained recognition as a leading tool for understanding protein folding, structure, and function. *De novo* computational design coupled with directed evolution can yield catalysts with novel functions and improved catalytic rate enhancement of enzymes.[1-2]

Minimalist protein design that focuses on the bare minimum requirements to achieve activity presents several important advantages. By utilizing basic physico-chemical properties and strategic placing of only few highly active residues one can feasibly sample a very large variety of possible catalysts. In more general terms minimalist approach looks for the mere *possibility of catalysis*, rather than trying to identify the *most active catalyst possible*. Even very basic designs that utilize a single residue introduced into non-enzymatic proteins or peptide bundles are surprisingly active. No complex calculations need to be set up and even a beginner can master this technique in a very short time. An enzyme nicknamed AlleyCatE is an allosterically regulated catalyst of ester hydrolysis that was generated by introducing a single histidine residue into a non-enzymatic protein calmodulin (CaM). The catalytic efficiency of the resulting enzyme is higher than that of any other rationally designed p-nitrophenylesterase and is on par with some catalytic antibodies.

DIRECTED EVOLUTION FOR THE DESIGN OF NEW CATALYSTS

By

Tiffany Trannelle Dunston
B.S. Syracuse University 2012
M.Phil. Syracuse University 2014

Dissertation

Submitted in partial fulfillment of the requirements for the degree
of Doctor of Philosophy in Chemistry

Syracuse University

May 2017

Copyright © Tiffany Trannelle Dunston 2017
ALL RIGHTS RESERVED

For my mother and grandmother.

Grandmom you shaped me into the woman I am today, I love you.
May your sweet soul rest in peace

For my father Edward Blount.
Like A Rock

ACKNOWLEDGEMENTS

I would like to start by thanking God for all the many blessings he has given me. I am extremely thankful and blessed to have a great family. You all are such a great support system and I could not ask for a better family. A special thanks to my grandmother and my mother for molding me into the scientist I am today.

I would like to thank my awesome graduate advisor Dr. Ivan Korendovych for giving me a great opportunity to do brilliant research in your lab. Your constant guidance, support, and wisdom is greatly appreciated. I would like to thank everyone in the Korendovych lab for your friendship and help and a warm extended thanks to Dr. Olesia Moroz, Dr. Yurii Moroz, for your guidance and kindness. I would like to thank all of my college professors from undergraduate to graduate school for teaching me great science.

Lastly, I would like to thank my friends for your constant encouragement and love. I would not be here today without all of you.

TABLE OF CONTENTS

1. Introduction.. 1

1.1 Enzymes.. 1

1.1.1 Properties of Enzymes...1

1.1.2 Structure and Function of Enzymes...3

1.2 De *novo* Protein Design...6

1.2.1 Computational Enzyme Design...7

1.2.2 Directed Evolution...8

1.2.3 Oligo Degeneracy via Saturation Mutagenesis....................................11

1.2.4 Protein Expression and Purification Protocols....................................13

1.2.5 Metalloprotein Design..17

1.3 Summary...17

2. Directed Evolution of an evolvable catalyst for an unnatural reaction...19

2.1 Introduction...19

2.2 De novo designing an artificial Kemp catalyst...............................20

2.2.1 Design Strategy……………………………………………………………………20

2.2.2 Michaelis-Menten Kinetics…………………………………………………26

2.2.3 Directed Evolution of CaM F92E 7th R for Enhanced Activity…….27

2.2.4 pK_a influence on catalytic residue………………………………………32

2.2.5 Calcium Dependence for Enzymatic Activation……………………….38

2.3 Conclusions…………………………………………………………………41

Appendix Chapter 2…………………………………………………………42

3. Site directed Mutagenesis of a non-enzymatic protein for a natural reaction……………………………………………………………………………47

3.1 Introduction………………………………………………………………47

3.2 Design and characterization of a robust efficient esterase……………………………………………………………………………50

3.2.1 Site-directed mutagenesis: Initial computational strategy………….50

3.2.2 Crude Cell Lysate Screening……………………………………………53

3.2.3 Exploring substrate recognition and stereospecficity of 4-nitro-2-phenyl propanoate and para-nitro phenyl acetate………………………………55

3.2.4 Protein Folding: Conformational Stability…………………………………60

3.2.5 MALDI-TOF for Capturing Acylated Intermediate..........................62

3.2.6 Assessing k_{cat} and product inhibition...64

3.2.7 Calcium Allosteric Regulation Assay..67

3.2.8 Characterization of other Full Length Esterases identified in the Saturation Mutagenesis work...69

3.3 Conclusion...72

Appendix Chapter 3...74

4. Small robust libraries in a thermostable enzyme for unnatural and natural reaction...79

4.1 Introduction..79

4.2 Library Design and Characterization of KE07 eliminase............80

4.2.1 Structural features for identifying important active site residues...80

4.2.2 Small robust PCR based libraries with a reduced amino acid alphabet...82

4.2.3 Crude Screening for the Identification of active hits.....................84

4.2.4 Kinetic Characterization of KE07 variants……………………………………86

4.3 Exploring tHisF for esterase activity……………………………………………90

4.3.1 Structural insight for exploring enzyme promiscuity……………………90

4.3.2 Kinetic characterization pNPB hydrolysis……………………………………92

. 4.3.3 Secondary structure analysis of tHisF variants………………………………97

. **4.4 Conclusion………………………………………………………………………………98**

Appendix Chapter 4………………………………………………………………………99

5. Experimental Procedures……………………………………………………………105

5.1 Materials……………………………………………………………………………………105

5.1.2 Protocol Preparation………………………………………………………………106

5.2 Computational Modeling……………………………………………………………108

5.2.1 Kemp eliminase F92E Design……………………………………………………108

5.2.2 Esterases…………………………………………………………………………………109

5.3 DNA Cloning/ Mutagenesis…………………………………………………………110

5.3.1 Ligated Independent Cloning……………………………………………………111

5.3.2 Screening Procedure for Kemp elimination…………………………………111

5.3.3 Crude cell Lysate Screening and Directed Evolution……………………112

5.4 Protein Expression and Purification………………………………………113

5.4.1 Esterases and Kemp eliminases…………………………………………113

5.4.2 KE07 10/11G variants……………………………………………………113

5.4.3 tHisF variants………………………………………………………………1114

5.4.4 Expression and Purification of Isotopically Labeled Proteins…………115

5.4.5 Affinity tag removal………………………………………………………116

5.5 Kinetic Assays………………………………………………………………116

5.5.1 Kemp Assays………………………………………………………………116

5.5.2 Esterase Hydrolytic Platereader Assay…………………………………117

5.5.3 Esterase UV-Vis Assay……………………………………………………118

5.5.4 pH Profiles…………………………………………………………………119

5.5.5 Calcium Dependent Assays………………………………………………119

5.6 Circular Dichroism Spectroscopy…………………………………………120

5.7 NMR Spectroscopy……………………………………………………………………121

5.8 Mass Spectrometry……………………………………………………………………122

5.9 Sequences………………………………………………………………………………123

References…………………………………………………………………………………124

LIST OF FIGURES

Figure 1.1 - *Reaction scheme depicting how enzymes overcome an energy barrier by lowering the activation energy of the transition state. Orange trace represents an uncatalyzed chemical reaction. The purple trace represents a catalyzed reaction with a smaller activation barrier to allow the reaction to proceed at a faster rate*..2

Figure 1.2 - *Lock and key model. Enzyme (purple) binds substrate (red) with the identical shape to fit in the active site (white space) to form enzyme-substrate complex to carry out catalysis*...5

Figure 1.3 - *Induced-fit model. Binding of substrate (blue) to the active site of the enzyme (white space) triggers a conformation change where the active site residues become more flexible to accommodate substrate*....................................5

Figure 1.4 - *Flow scheme for the process of directed evolution. Evolution starts with a gene of interest which is randomized through mutagenesis to generate a* *large diverse library of variants (color coded). Those variants are expressed, screened, and isolated via high throughput screening method. The desired variant serves as a new template for subsequent iterative rounds of mutagenesis*……………………………………………………………………………………………10

Figure 1.5 *- Site directed mutatgenesis process. Primers containing the codon for a histidine CAC complements the parent plasmid containing codon for methionine ATG and is extended in PCR to generate a new strand of DNA. The newly synthesized DNA strand represents the new mutated DNA. Parent DNA is destroyed by DpnI digestion and the resulting mutated gene is transformed into E.coli to produce colonies with mutated gene*…………………………………………12

Figure 1.6 *- Nickel NTA affinity chromatography coordination. NTA agarose tightly chelates nickel metal ion. The polyhistidine tag in the protein sequence coordinates to the nickel displaying high binding affinity. His tagged proteins bind to the column in this fashion while nonspecific proteins without the tag flow through*……………………………………………………………………………………………16

Figure 2.1 *- Unbound state of calmodulin (left). Bound state of calmodulin with calcium ions as red spheres, PDB 3CLN*…………………………………………………21

Figure 2.2 - *Kemp elimination reaction.*

5-Nitrobenzisoxazole as the substrate..22

Figure 2.3 - *C-terminal domain of CaM with potential active site catalytic residues (yellow spheres) for accomodating carboxylates...23*

Figure 2.4 - *Van der Waals energies of Asp (blue) or Glu (pink) as lowest rotamers at each hydrophobic residue 24*

Figure 2.5 - *Left: Examination of the Glu residues in the identified positions docking of the substrate. Right: Transition state docking based on a search of a superrotamer library for the transition state..25*

Figure 2.6 - *Enzyme's behavior demonstrated by the Michaelis -Menten curve. Reaction velocity as a function of substrate concentration 27*

Figure 2.7 - *Kemp elimination activities of S81R CaM F92E 7th R and of S81R CaM F92E 7th R C-terminal mutant with affinity tag. Conditions: 20 mM HEPES (pH 7.0); 100 mM NaCl; 10 mM CaCl2; substrate concentration 0.00012-0.00096 M; 0.2 μM S81R CaM F92E 7th R; 0.18 μM S81R CaM F92E 7th R C-terminal; the catalytic efficiencies obtained from fitting the linear portion of* *the curves are k_{cat}/K_M 882 ± 1.6 $M^{-1}s^{-1}$, k_{cat}/K_M 447 ± 6.5 $M^{-1}s^{-1}$ respectively*..........29

Figure 2.8 *- Kemp Elimination activity of K94M CaM F92E 7th R mutant with affinity tag. Conditions: 20 mM HEPES (pH 7.0); 100 mM NaCl; 10 mM $CaCl_2$; substrate concentration 0.12-0.96 mM; 0.2 μM K94M CaM F92E 7th R ; the catalytic efficiencies obtained from fitting the linear portion of the curves are (k_{cat}/K_M = 834 ± 20 $M^{-1}s^{-1}$)*..........30

Figure 2.9 *- Kemp Elimination activity of S81R & K94M CaM F92E 7th R mutant without affinity tag. Conditions: 20 mM HEPES (pH 7.0); 100 mM NaCl; 10 mM $CaCl_2$; the catalytic efficiencies obtained from fitting the linear portion of the curves are k_{cat}/K_M = 721 ± 5.3 $M^{-1}s^{-1}$ and 504 ± 5 $M^{-1}s^{-1}$ respectively*..........31

Figure 2.10 *- pH profiles of original scaffold CaM F92E (left) and CaM F92E 7thR (right). Conditions: pH 4.5-8.5, 150 μM substrate*..........32

Figure 2.11 *- pH profile of S81R with and without affinity tag. Conditions: pH 4.5-9.0; 150 μM substrate; 0.20 μM S81R*..........33

Figure 2.12 *- pH profile of K94M with and without affinity tag.*

Conditions: pH 4.5-9.0; 150 μM substrate; 0.20 μM K94M.................................34

Figure 2.13 *- pH profile of V142L with affinity tag. Conditions: pH 4.5-9.0; 150 μM substrate; 0.20 μM K94M...35*

Figure 2.14 *- pH profile of D133N with affinity tag. Conditions: pH 4.5-9.0; 150 μM substrate; 0.20 μM D133...36*

Figure 2.15 *- pH profile of T110R with affinity tag. Conditions: pH 4.5-9.0; 150 μM substrate; 0.20 μM T110R...37*

Table 1 *- Kinetic Data summary of ep-PCR variants evolved from CaM F92E 7th R..38*

Figure 2.16 *- Dependence of the activity of K94M on* $[Ca^{+2}]$ *at pH 7.0 in 20 mM HEPES; 100 mM NaCl; substrate concentration 100 μM; protein concentration 0.20 μM.* $(k_{cat}/K_M)_{max} = 728 \pm 18\ M^{-1}s^{-1}$, $M_{50} = 21.8 \pm 1.7\ \mu M$, $h = 1.6 \pm 0.20$*..39*

Figure 2.17 *- Dependence of the activity of D133N on* $[Ca^{+2}]$ *at pH 7.0 in 20 mM HEPES; 100 mM NaCl; substrate concentration 100 μM; protein concentration 0.20 μM.* $(k_{cat}/K_M)_{max} = 720 \pm 16\ M^{-1}s^{-1}$, $M_{50} = 23.8 \pm 1.7\ \mu M$, $h = 1.5 \pm 0.15$*..40*

Figure S.01 *- Kemp Elimination activity of D131G CaM F92E 7th R mutant fit to Michaelis-Menten equation pH 7.0, 0.09 μM protein..42*

Figure S.02 *Kemp Elimination activity of N111D CaM F92E 7th R mutant fit to Michaelis-Menten equation pH 7.0, 0.10 μM protein...42*

Figure S.03 *Kemp Elimination activity of V142L CaM F92E 7th R mutant fit to Michaelis-Menten equation pH 7.0, 0.21 μM protein..43*

Figure S.04 *Kemp Elimination activity of K94M CaM F92E 7th R mutant fit to Michaelis-Menten equation pH 8.0, 0.20 μM protein..43*

Figure S.05 *Kemp Elimination activity of D80H CaM F92E 7th R mutant fit to Michaelis Menten equation pH 7.0, 0.20 μM protein...44*

Figure S.06 *Kemp Elimination activity of G96C CaM F92E 7th R mutant fit to Michaelis Menten equation pH 7.0, 0.28 μM protein...44*

Figure S.07 *- Kemp Elimination activity of D133N CaM F92E 7th R mutant fit to Michaelis-Menten equation pH 7.0 (left), pH 8.0 (right) 0.20 μM protein...45*

Figure S.08 *- Kemp Elimination activity of T110R CaM F92E 7th R* *mutant fit to Michaelis-Menten equation pH 7.0 (left), pH 8.0 (right) 0.21 μM protein*……………………………………………………………………………………45

Figure S.09 *- Kemp Elimination activity of I107V CaM F92E 7^{th} R mutant fit to Michaelis-Menten equation*………………………………………………………………46

Figure S.010 ***-*** *10% SDS-PAGE CaM F92E 7^{th}R mutants stained with coomassie blue. S81R and K94M have tags removed by TEV protease* ……………………………46

Figure 3.1 *- Hydrolysis of p-nitrophenyl (2-phenyl)-propanoate (pNPP)*……………………………………………………………………………………49

Figure 3.2 *- Histidine mediated hydrolysis of p-nitrophenyl esters. Reaction demonstrating "classic" burst phase kinetics*………………………………………………50

Figure 3.3 *Computational methodology employed for identifying desired variant for ester hydrolysis reaction. Potential sites for introduction of histidine are outlined in orange*……………………………………………………………………52

Figure 3.4 *pNPP hydrolysis by crude cell lysates of E. coli: BL21 (DE3) pLysS cells (black triangles); E. coli BL21(DE3) pLysS cells containing a plasmid encoding c-CaM gene (green diamonds); E. coli BL21(DE3) pLysS cells* containing a plasmid encoding AlleyCatE (red circles); and E. coli BL21(DE3) pLysS cells containing a plasmid encoding full length CaM M144H (blue squares). Conditions: 20 mM HEPES (pH 8.0); 100 mM NaCl; 10 mM CaCl2; 0.2% Triton X-100; 25 μM pNPP. Absorbance was monitored at 405 nm..........54

Figure 3.5 *R and S isomers of pNPP*..........55

Table 3.1 *- Kinetic parameters for AlleyCatE and AlleyCatE2*..........56

Figure 3.6a *Analysis of the kinetic data for hydrolysis of S-pNPP catalyzed by AlleyCatE. Conditions: 20 mM HEPES (pH 7.5); 100 mM NaCl; 10 mM* $CaCl_2$; *2.0 μM AlleyCatE*..........57

Figure 3.6b *Analysis of the kinetic data for hydrolysis of S-pNPP catalyzed by AlleyCatE2. Conditions: 20 mM HEPES (pH 7.5); 100 mM NaCl; 10 mM* $CaCl_2$; *2.0 μM AlleyCatE2*..........57

Figure 3.7a *- Analysis of the kinetic data for hydrolysis of R-pNPP catalyzed by AlleyCatE. Conditions: 20 mM HEPES (pH 7.5); 100 mM NaCl; 10 mM* $CaCl_2$; *2.0 μM AlleyCatE*..........58

Figure 3.7b - *Analysis of the kinetic data for hydrolysis of R-pNPP catalyzed by AlleyCatE2. Conditions: 20 mM HEPES (pH 7.5); 100 mM NaCl; 10 mM $CaCl_2$; 2.0 μM AlleyCatE2*......58

Figure 3.8a - *Analysis of the kinetic data for hydrolysis of pNPA catalyzed by AlleyCatE. Conditions: 20 mM HEPES (pH 7.5); 100 mM NaCl; 10 mM $CaCl_2$; 6.0 μM AlleyCatE*......59

Figure 3.8b - *Analysis of the kinetic data for hydrolysis of pNPA catalyzed by AlleyCatE2. Conditions: 20 mM HEPES (pH 7.5); 100 mM NaCl; 10 mM $CaCl_2$; 6.0 μM AlleyCatE2*......59

Figure 3.9 - *CD spectrum of AlleyCatE and CaM-M144H. Conditions: 25.0 μM protein, 4 mM HEPES (pH 7.5), 2 mM CaCl2 and 30 mM NaCl*......60

Figure 3.10 - *Chemical denaturation (guanidium hydrochloride GdmCl) profiles of c-CaM (blue circles) and AlleyCatE (red circles)*......61

Table 3.2 - *Thermodynamic parameters obtained from denaturation studies*......62

Figure 3.11 - *MALDI-TOF spectra of A: free AlleyCat E, B: AlleyCatE with 0.5 equiv.pNPP, C: free AleyCatE2, D: AlleyCatE2 with 0.5 equiv pNPP* 63

Figure 3.12 *MALDI-TOF spectra of c-CaM in the presence of 2.5 equiv R-pNPP*..64

Figure 3.13 *MALDI-TOF spectra of AlleyCatE at taken at different time points . Top: AlletCatE in the presence of 8 equivalents pNPP. The three peaks (zoomed in) shows AlleyCatE single, double, and triply aclyated. Bottom: Full conversion of pNPP after 48 hours. The one peak shows AlleyCatE acylated four times*...65

Figure 3.14 - *Substrate turnover by AlleyCatE. Conditions: 2.0 μM AlleyCatE, 60 μM R-pNPP in 20 mM HEPES, 100 mM NaCl, 10 mM $CaCl_2$, pH 7.5*................66

Figure 3.15 - *The photometric traces (at 405 nm) of R-pNPP hydrolysis catalyzed by AlleyCatE in the absence of product (black), in the presence of 1 equiv. of p-nitrophenol (light blue), in the presence of 5 equiv. of p-nitrophenol (dark blue), and in the presence of 5 equiv. of (R)-2-phenylpropionic acid (red).*

Conditions: 2.0 μM AlleyCatE, 60 μM R-pNPP in 20 mM HEPES, 100 mM NaCl, 10 mM $CaCl_2$, pH 7.5.........67

Figure 3.16 *- A) Hydrolysis of R-pNPP (50 μM) catalyzed by AlleyCatE (2 μM) in the presence (blue) and in the absence (red) of 10 mM CaCl2. B) Hydrolysis of R-pNPP (50 μM) catalyzed by AlleyCatE (2 μM) in the presence (blue) and in the absence (red) of 10 mM CaCl2. Conditions: 20 mM HEPES, 100 mM NaCl, pH 7.5*.........68

Figure 3.17 *- Left: pNPP burst phase kinetics of A128T, V108G, and wild type V108V (M144H) at pH 7.5. Right: pNPP slow phase reaction of A128T, V108G, and wild type V108V (M144H) at pH 7.5, 4.5 μM protein*.........70

Figure 3.18 *- Left: pNPA burst phase kinetics of A128T, V108G, and wild type V108V (M144H) at pH 7.5. Right: pNPA slow phase reaction of A128T, V108G, and wild type V108V (M144H) at pH 7.5, 4.5 μM protein*.........70

Figure 3.19 *- pNPP slow phase kinetics of A128T, V108G, and wild type V108V (M144H), A88G, and A88T at pH 7.5, 2.0 μM protein*.........71

Figure 3.20 *- pNPP kinetics V108GA128T and wild type V108V (M144H at pH 7.5, 2.0 μM protein*.........72

Figure S.01 - *10% SDS-PAGE of cCaM, AlleyCatE, and AlleyCatE2 proteins in 1X MES buffer* 74

Figure S.02 - *pH profiles of burst phase rate of AlleyCatE and AlleyCatE2. AlleyCatE shows a p*K_a *of 6.8 ± 0.07 with a* k_{cat}/K_{Mmax} *of 160± 5.2, AlleyCatE2 p*K_a *of 6.9 ± 0.14 with a* k_{cat}/K_{Mmax} *of 220± 14. Conditions: pH 4.5-9.5; 20 μM of substrate; 2.0 μM proteins. * pH's 4.5-5.5 were excluded*..74

Figure S.03 - *Michaelis-Menten plot for hydrolysis of paraoxon catalyzed by AlleyCatE (left) and AlleyCatE2 (right) variant Conditions: 20 mM HEPES (pH 7.5); 100 mM NaCl; 10 mM* $CaCl_2$*; 10.0 μM proteins.* k_{cat}/K_M *for proteins are 0.10 and 0.16 (*$M^{-1}s^{-1}$*) respectively*..75

Figure S.04 - *Michaelis-Menten plot for hydrolysis of pNPA catalyzed by M109W AlleyCatE variant.* k_{cat}/K_M *of 9 ± 0.5* $M^{-1}s^{-1}$..76

Figure S.05 - *Left: PCR products on 0.8% agarose gel on positions 92NNK (lanes 1-3), 109NNK (lanes 4-6), 124NNK (lanes 7-9). Right: 127NNK (lanes 1-3), 145NNK (lanes 4-6), and 148NNK (lanes 7-9)*...76

Figure S.06 - *Left: 50 μM and 25 μM coumarin no base emission scan excited at 310 nm with 10 nm slit. Substrate emits weakly at 383 nm. Right: 50 μM coumarin with 30 mM NaOH emission from 330-600 nm*.......77

Figure S.07 - *Fluorescent platereader kinetic data of coumarin hydrolysis by AlleyCatE and AlleyCatE2 excited at 310 nm, emitted at 466nm for 1 hr. Final substrate concentration 750 μM, 10 μM protein concentration at pH 7.5. AlleyCatE or AlleyCatE2 does not hydrolyze coumarin*.......77

Figure S.08 - *Fluorescent plate reader emission spectrum at 750 μM coumarin substrate with 10 μM protein concentration pH 7.5 for 1 hr. Plate excited at 310 nm and emitted from 330-700 nm*.......78

Figure 4.1 *(a) In the original crystal structure of KE07, the Glu101-Lys222 distance is 2.8 Å. (b) With the additional polar mutation D7 in close proximity, it disrupts the Glu101-Lys222 interaction and the distance increases, making Glu101 a more reactive base, PDB code: 1THF* 81

Figure 4.2 *Theozyme model for Kemp elimination reaction. A: Kemp elimination displaying transition state. B: Designed carboxylate and histidine* *carboxylate motif incorporating surrounding active site residues; pi stacking residues (tryptophan) and H-bond donor lysine for stabilizing transition state. Right: Cartoon depiction demonstrating algorithms applied to theozyme model..81*

Figure 4.3 *(left) Secondary structures α helices (cyan) and β sheets (magenta) of Library A with positions to be mutated represented by sticks A130 and Y128 and a sphere G171. (right) Secondary structures α helices (cyan) and β sheets (magenta) of Library B with positions to be mutated represented by sticks K222 and D7. PBD: 3iiv..84*

Figure 4.4 *- Sanger DNA sequencing of Library A diversity at positions Y128 (ATT), A130 (GTT), and G172 (GCC)..85*

Figure 4.5 *(a) Confirmation screen of combined library A and B hits with rows A-B blank, C1-C6= WT, C7-C12 = libA , D1-D6 = libA , D7-D1= G6, E1-E6= libA, E7-E12= libB, F1-F6= libB, F7-F12 libB (b,c) Confirmation screen of library C. (b) WT = G9-G12, others are confirmed variants from intial screens. (c) WT = C4-C7, others are confirmed variants from initial screen.................................85*

Figure 4.6 - *Kemp Elimination activity of KE07 10/11G, G171S, and K222M. Conditions: 25 mM HEPES (pH 7.0); 100 mM NaCl, 0.25 μM protein. Data were fit to Michaelis-Menten linear regime* 87

Figure 4.7 - - *Kemp Elimination activity of KE07 10/11G and G171A. Conditions: 25 mM HEPES (pH 7.0); 100 mM NaCl, 1.0 μM protein. Data were fit to Michaelis-Menten linear regime* 87

Figure 4.8- *Elimination activity of KE07 10/11G, A130S, and I129V. Conditions: 25 mM HEPES (pH 7.0); 100 mM NaCl, 0.25 μM protein. Data were fit to Michaelis-Menten linear regime*......88

Figure 4.9 - *Kemp Elimination activity of KE07 10/11G, A9G, and A130T. Conditions: 25 mM HEPES (pH 7.0); 100 mM NaCl, 0.25 μM protein. Data were fit to Michaelis-Menten linear regime*......89

Figure 4.10 - Kemp *Elimination activity of KE07 10/11G and I129VA130S. Conditions: 25 mM HEPES (pH 7.0); 100 mM NaCl, 0.25 μM protein. Data were fit to Michaelis-Menten linear regime*......89

Figure 4.11 - *Natural reaction of bienzyme complex tHisH-tHisF* 91

Figure 4.12 - *Crystal structure of tHisF (PDF code: 1THF) with α helices (cyan) and β sheets (magenta) and buffer phosphate as spheres. Asp11 and Asp130 are important catalytic residues. Histidine (green) were mutated to double glutamines or alanines..92*

Figure 4.13 - *Solubility of pNPB at various substrate concentrations in the presence of 1.5% (top) and 10% (bottom) acetonitrile...93*

Figure 4.14 - *Michaelis-Menten plot of WT tHisF. The k_{cat}/K_M value obtained from linear fit (not shown) was 0.10 ± 0.02 at pH 7.4, 18 μM protein.................94*

Figure 4.15 - *Michaelis-Menten plot of H84QH228Q (QQ) tHisF. The k_{cat}/K_M value obtained from linear fit (not shown) was 0.11 ± 0.01 at pH 7.4, 18 μM protein...94*

Figure 4.16 - *Michaelis-Menten plot of H84AH228A (AA) tHisF. The k_{cat}/K_M value obtained from linear fit (not shown) was 0.11 ± 0.01 at pH 7.4, 18 μM protein...95*

Figure 4.17 - *Michaelis-Menten plot of H228A (228A) tHisF. The k_{cat}/K_M value obtained from linear fit (not shown) was 0.11 ± .01 at pH 7.4, 18 μM protein....95*

Figure 4.18 *- Michaelis-Menten plots of c-CaM and calmodulin no lysines (caM no K) with* k_{cat}/K_M *values of 0.13 ± 0.03 and 0.5 ± 0.1 respectively deduced from linear fit (not shown) at pH 7.4, 6 μM protein……………………………………96*

Figure 4.19 *CD on all variants reveals the α-helical structure at pH 7.4 with 18 μM…………………………………………………………………………………………97*

Figure S.01 *- Left: Gel A megaprimer product from library A. Right: Gel B PCR extension products generated from library A megaprimer at different annealing temperatures (Lanes 1-5) on a 1% agarose gel………………………………99*

Figure S.02 *- Left: Gel A megaprimer product from library B. Right: Gel B PCR extension products generated from library B megaprimer at different annealing temperatures (Lanes 1-5) on a 1% agarose gel………………………………99*

Figure S.03 *- Left: Gel of megaprimer product from library C. Right: Gel of PCR extension products generated from library C megaprimer at different annealing temperatures (Lanes 1-5) on a 1% agarose gel. Order: closest to ladder 1-5 (KE07 10/11G template), farthest 1-5 (K222M template)……………………………………………………………100*

Figure S.04 - *Left: PCR megaprimer for library C at different primer concentrations. Right: PCR extension using megaprimer 7, 11, and 13 from the left gel*..........100

Figure S.05 - *PCR of D7NTT position on KE07 10/11G template.*
** K222M and I129VA130S templates were done but did not work*..........101

Figure S.06 - *0.8% agarose gel on PCR products. Shown in lane 1: H84AH228A (AA), lane 2: H84QH228Q (QQ), lane 3: H84QH228Q (QQ), lane 4: H84QH228Q (QQ)*..........101

Figure S.07 - *15% SDS-PAGE of G172A and WT KE07 10/11 with different growth conditions (ZYM and IPTG 37C and 18C). Lane 6: soluble G172A ZYM 18C, Lane 10: soluble G172A IPTG 18C, Lane 14: soluble WT IPTG 18C. *other lanes are expression and debris pellet of the proteins*..........102

Figure S.08 - *15% SDS-PAGE: Left: 5 µg of G172A (lane 1) and WT (lane 2). Middle: I129V (lane 1), A130S (lane 2). Right: (lane 1-4) I129VA130S, A130T, WT, A9G*..........102

Figure S.09 - *Left: Fractions from G25 Desalting of tHisF. Right: Pure tHisF fractions collected after Q Sepharose purification...103*

Figure S.010 - *Left: Fractions from Q Speharose of tHisF (3-7),*

QQ (lanes 3-5 at 200 mM NaCl) (lanes 1-6 at 250 mM NaCl).

Right: Fractions from Q Sepharose of AA (lanes 3-5 at 200 mM NaCl,

lanes 1-5 at 250 mM NaCl)...103

Figure S.011 - *Pure proteins buffer exchanged pH 7.4. Order from ladder:*

tHisF, QQ, AA, and 228A..104

LIST OF ABBREVIATIONS

AlleyCatE	Allosterically controlled esterase catalyst
Asp	Aspartate
Kemp	1,2-Benzisoxazole
Ca^{+2}	Calcium
CaM	Calmodulin
CD	Circular Dichroism
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
ep-PCR	error prone Polymerase Chain Reaction
E.*coli*	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
Glu	Glutamate
GdmCl	Guanidine Hydrochloride
His	Histidine
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
IPTG	Isopropyl β-D-1-thiogalactopyranoside

LB	Luria Bertani
MALDI-TOF	Matrix assisted laser desorption/ionization-time of flight
MM	Michaelis-Menten
MES	2-(N-morpholino)ethanesulfonic acid
NTA	Nitrilotriacetic acid
NMR	Nuclear magnetic resonance
pNPA	para-nitrophenyl acetate
pNPB	para-nitrophenyl butyrate
pNP	para-nitrophenol
pNPP	para-nitrophenyl-(2-phenyl)-propanoate
PCR	Polymerase Chain Reaction
SDS-PAGE	Sodium Dodecyl sulfate polyacrylamide gel electrophoresis
TEV	Tobacco Etch Virus
TRIS	tris(hydroxymethyl)aminomethane
TAPS	N-Tris (hydroxymethyl) aminomethane
UV-Vis	Ultraviolet-visible spectroscopy

1. Introduction

1.1 Enzymes

1.1.1 Properties of Enzymes

Enzymes are proteins that are essential for all human life. The majority of enzymes are proteins. They are responsible for speeding up and regulating biochemical reactions that takes place in the human body.[3-5]

During the catalyzed process, enzymes are not consumed or altered. In fact, they facilitate reactions by lowering the activation energy, which is the minimum energy required to drive the reaction to completion. As shown in Fig 1.1, enzyme catalyzed reactions have a lower activation energy than an uncatalyzed reaction.[3-4] The Gibbs free energy remain constant during the course of the reaction for both reactants and products, however the activation energy between the transition state and the reactants is changing.

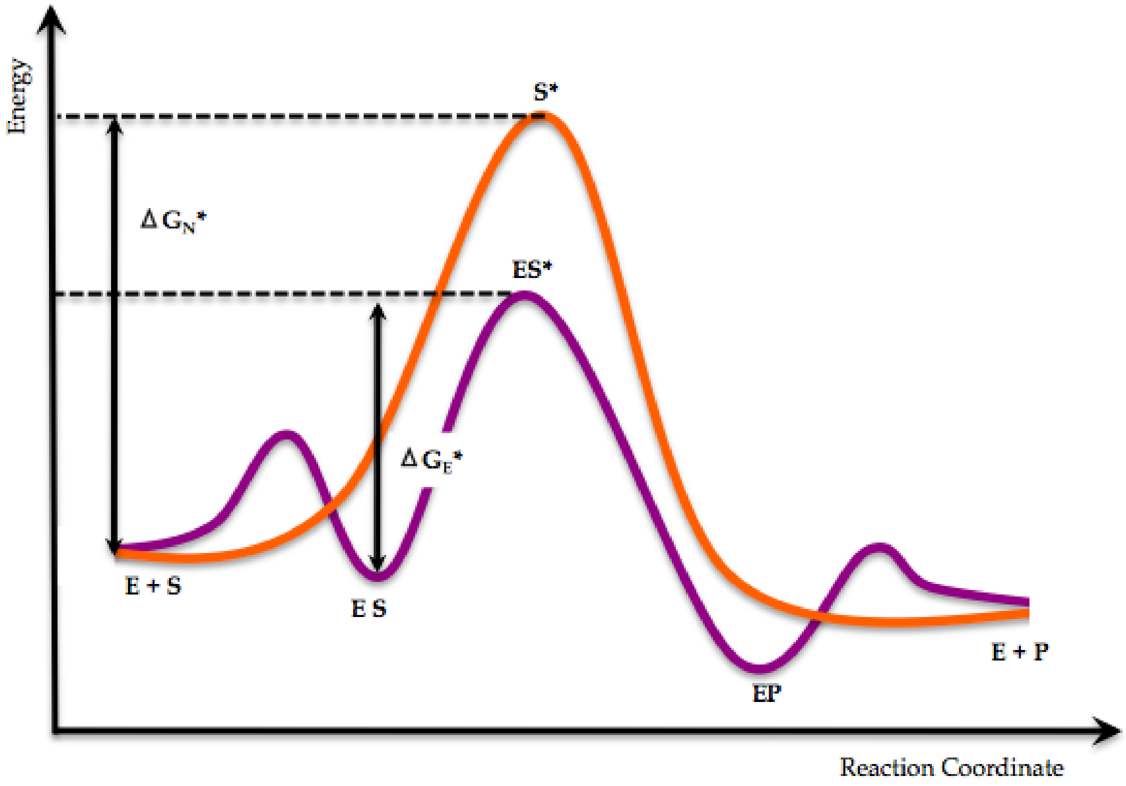


Figure 1.1 Reaction scheme depicting how enzymes overcome an energy barrier by lowering the activation energy of the transition state. Orange trace represents an uncatalyzed chemical reaction. The purple trace represents a catalyzed reaction with a smaller activation barrier to allow the reaction to proceed at a faster rate.[6]

Enzymes are highly versatile and specific molecules. They can easily identify and target a specific substrate and carry out a single reaction of a set of similar reactions efficiently. Catalysis occurs inside the enzyme's active site. The active site contains the catalytic residues of the enzymes, thus is responsible for stabilizing the transition state.[7-8] By creating such microenvironment, substrate molecules are oriented and bought in close proximity for the reaction to take place and decreases the likelihood of unwanted side reactions.

Some but not all enzymes are stereospecific. The enzyme can only recognize reactants' that are oriented in a specific conformation and can only catalyzes its reaction if the stereochemistry is correct. For example, chymotrypsin is an enzyme that hydrolyzes proteins using a catalytic triad motif.[9] It selectively targets the peptide bond next to a phenylalanine residue and that phenylalanine must have the S-configuration [9] for the reaction to occur. If phenylalanine had the R-configuration, the active site will not accommodate that substrate and thus the reactions will not take place.

1.1.2 Structure and Function of Enzymes

Proteins are composed of a variety of amino acids that have distinct properties. Proteins are macromolecules that function based on their primary amino acid sequence, which in turn influences its 3-dimensional conformation. This sequence can be modified in a way that can enhance or disturb the protein's overall biological activity.

Proteins are natively folded into 3D shapes. Some proteins are thermally stable or only stable under certain conditions. The native structure can become perturbed if the protein is exposed to external oxidative stresses, high pH, high temperature, or pressure. [5, 10]These disturbances in the proteins structure can lead to impairment in the biological activity.

Catalysis in enzymes can be assisted by cofactors such as metals ions. In some cases, a protein's activity is only regulated in the presence of a metal cofactor such as zinc, and these catalyst are termed metalloenzymes.[11-14]

There are two models that describes enzymes specificity known as the lock and key model proposed by Emil Fischer and the induced fit proposed by Daniel Koshland (Figure 1.2-1.3). The lock and key model states that the enzymes active site is a perfect key hold for the substrate (the key).[3] The structure of neither the substrate nor enzyme change shape during the binding. The more accepted model is described by the induced fit theory, which states that the binding of a substrate to the protein's active site induces a change in the shape of the enzyme to enhance its activity.[3] It is the electrostatic interactions between molecules that is responsible for inducing the change in shape.

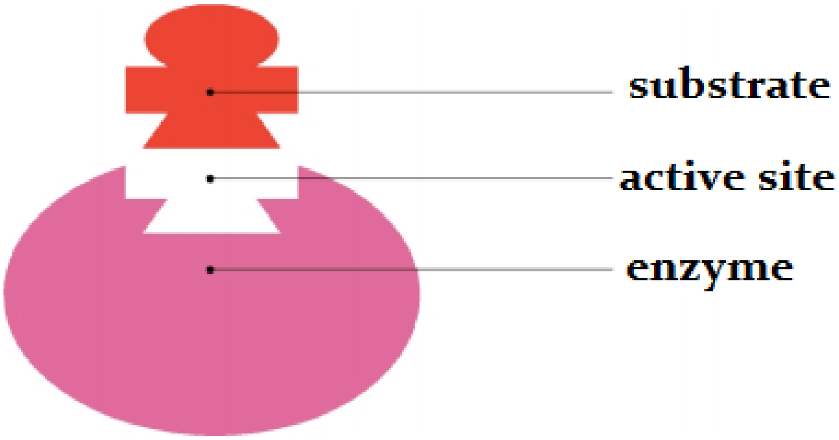

substrate
active site
enzyme

Figure 1.2 Lock and key model. Enzyme (purple) binds substrate (red) with the identical shape to fit in the active site (white space) to form enzyme-substrate complex to carry out catalysis.

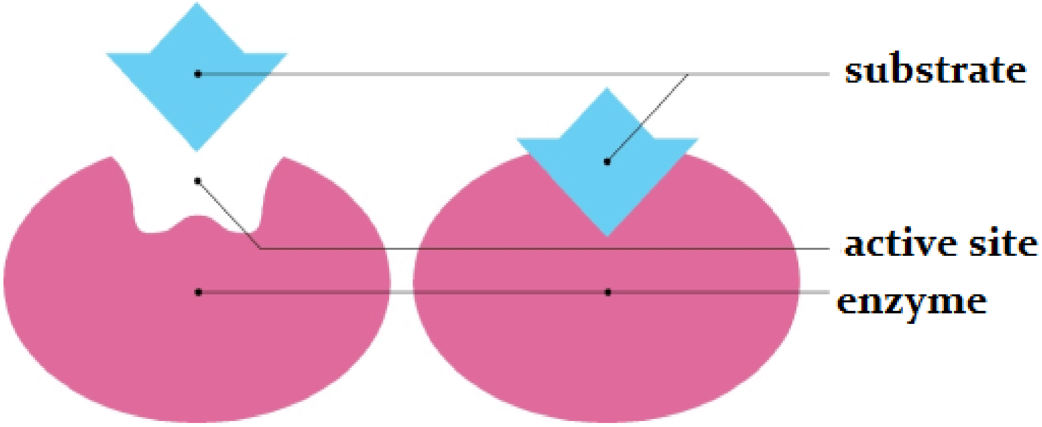

substrate
active site
enzyme

Figure 1.3 Induced-fit model. Binding of substrate (blue) to the active site of the enzyme (white space) triggers a conformation change where the active site residues become more flexible to accommodate substrate.

1.2 *De novo* Protein Design

Over the past few years, much research has been done to implement protein design strategies for the creation of novel proteins for biochemical reactions. Protein engineering has developed ways to produce proteins with a wide range of industrial and health related applications. More specifically, protein engineering is known to overcome the challenges and limitations of naturally occurring enzymes.[15] De novo protein design is a rational design of creating new protein molecules with new function.[16-17] Pre-existing protein scaffolds are utilized as starting points for evolution with the ultimate goal of designing new behavior and novel catalytic efficiencies.[18] Since there are many classes of enzymes that exists in nature, one can utilize the physical and chemical properties of those enzymes to reconstruct and design proteins beyond the limits of natural protein structures.[19]

Despite enormous functional diversity in different living organisms, there are many industrial synthetic pathways for which no enzymes are known. Thus, creating new functionality in proteins is an important challenge. Several methodologies have been developed to create new activities in proteins such as directed evolution and complex computational enzyme design.[20] The idea of creating new functional proteins opens up new questions that must be tested computationally and experimentally: How do we build into protein-structure space when there are few natural structures to guide us? To what uses can the resulting proteins be put? And what would this new design tell us about how natural proteins fold, function, and evolve? Most of these questions are answered using protein structure prediction and computational analysis.[21]

1.2.1 Computational Enzyme Design

Research has been done in investigating the principles of protein design strategies in addition to understanding the physiochemical properties of enzyme catalysis. Proteins can fold into well-defined structures and act as biocatalysts[22] for reactions, but for some the catalytic rate remains low. Although a success, enzyme optimization still remains a challenge due to enzyme size, substrate range, and other external factors that can affect an enzyme's ability to function properly. Computational enzyme design has become remarkably useful tool to overcome these limitations. *De novo* computational design coupled with directed evolution can yield catalysts with novel functions and improved catalytic rate enhancement of enzymes.[17, 23-25]

Computational strategies such as computer based algorithms[26] has paved the way for simplifying protein design models. Computational models are constructed for both protein folding prediction, protein design, and is used to make predictions about amino acid sequences for a desired target protein. When designing these models it is important to answer the following question: What are the underlying physicochemical properties that govern enzyme catalysis in terms of protein folding, function, and stability?[27] To computationally study some of these enzymatic properties, one can run a design calculation that incorporate specified amino acids. Once the amino acids are identified, interaction energy computations are done by evaluating the potential energy functions of the protein. This will thermodynamically assess the proteins most favorable interactions ultimately generating the lowest energy protein sequence.[28-32]

Over the course of evolution, it remains a challenge for nature to test all possible protein sequences or structures. To screen all 20 amino acid combinations in all positions of a 100 amino acid protein, this would yield 20^{100} sequence combinations. For a sequence this large, it is experimentally impossible to screen with laboratory methods that will allow one to explore full sequence diversity. To bypass this limitation, computationally one can utilize combinatorial algorithms such as RosettaDesign and Monte Carlo[31, 33] to do randomized searches in small sequence space of the protein to ultimately *de novo* design proteins with novel catalytic functions on par with naturally occurring enzymes.

1.2.2 Directed Evolution

The process of directed evolution mimics that of natural selection. It is a technique used in protein engineering aimed to produce enzymes with new functionalities.[34] Directed evolution can be used to target enzymatic characteristics such as pH, thermal stability, substrate specificity, and activity.[35-36] It is initiated by subjecting a gene of interest to several rounds of mutagenesis[37] via methods such as PCR, gene shuffling, and/or saturation mutagenesis.[38-39] These iterative rounds of mutagenesis creates a library of DNA mutants. Next, the variants are expressed, screened, and isolated for a desired activity. The isolated variants with higher enzymatic activity than wild-type protein will then serve as templates for further rounds of mutagenesis. The directed evolution process is outlines in Figure 1.4.

For any directed evolution experiment to be successful, it is essential to choose a library design strategy. There are many ranging from mutagenesis (random or semi-rational) and gene recombination.[40] Random mutagenesis randomly introduced mutations throughout the entire gene target. These mutations can be classified as point mutations, insertions, frame-shift, or deletions. Gene recombination or gene shuffling is another useful technique that exchange gene fragments from different parental DNA to form a chimera. Lastly, semi-rational is a combination of both random and site-directed mutagenesis where specific amino acid positions that are important in the protein's function are randomized to all 20 amino acids.[16, 40-41]

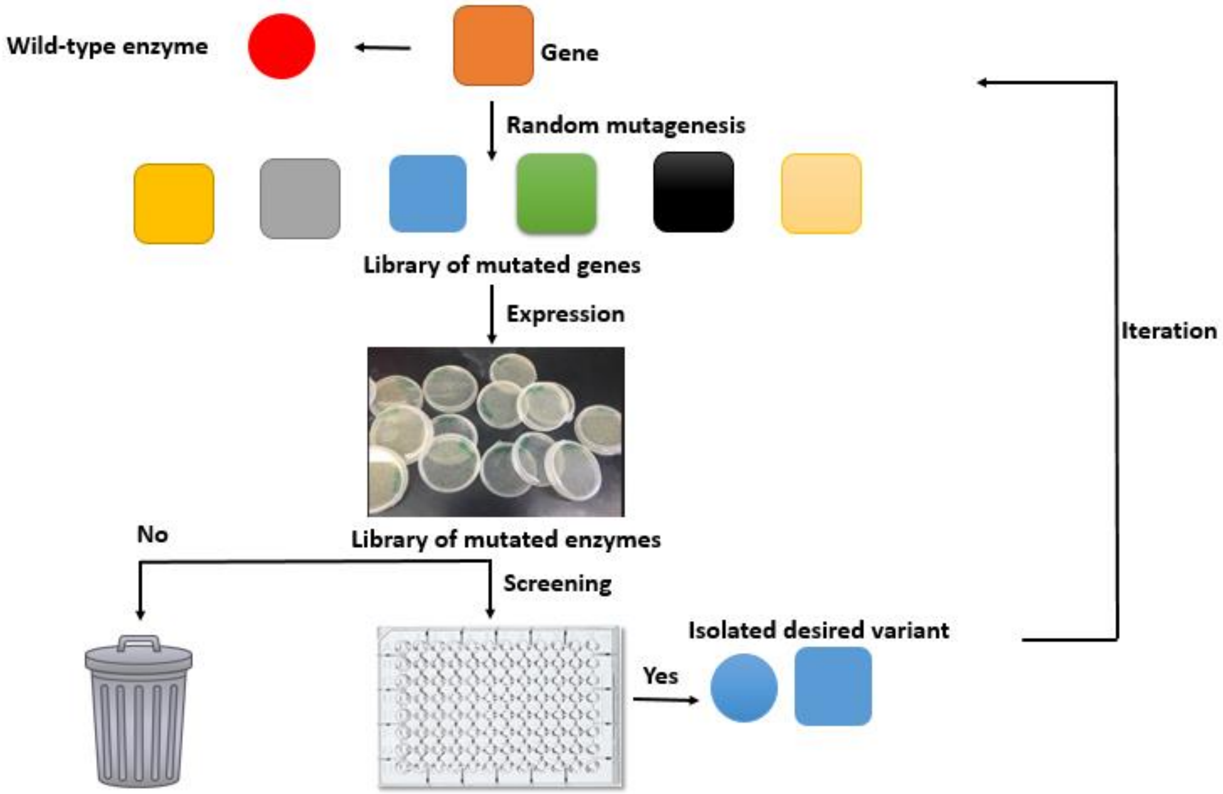

Wild-type enzyme
Gene
Random mutagenesis
Library of mutated genes
Expression
Library of mutated enzymes
No
Screening
Isolated desired variant
Yes
Iteration

Figure 1.4 Flow scheme for the process of directed evolution. Evolution starts with a gene of interest which is randomized through mutagenesis to generate a large diverse library of variants (color coded). Those variants are expressed, screened, and isolated via high throughput screening method. The desired variant serves as a new template for subsequent iterative rounds of mutagenesis.

1.2.3 Oligo Degeneracy via Saturation Mutagenesis

Library creation methodology is an important aspect for efficient mutagenesis.[42] This can be achieved through a process known as saturation mutagenesis. Saturation mutagenesis introduce site specific mutations in a gene of interest.[40] A primer binds complementary to the vector, contains the desired mutation, and is extended around the DNA plasmid *via* a technique known as PCR.[39, 43] Polymerase chain reaction allow for the production of several to billions of copies of the desired mutant gene (Figure 1.5).[39]

Creating high quality libraries are experimentally challenging. Many factors play a role in the resulting library such as temperature, G+C content, annealing, and primer quality.[44] To minimize the amount of failure one can encounter with the library, it is important to modify such conditions.

Randomizing positions with degenerate codons offers a great advantage for increasing library diversity. Degenerate primers are sometime used in saturation mutagenesis to provide more flexibility in the DNA sequence for generating all possible amino acid combinations at a specific position.[45] Such degenerate codons include the NNK motif, where N represents equimolar amounts of each nucleic base (A, T, C, and G) while K represents either guanine or thymine.[46] With this broad range of amino acid diversity, it will cover the full range of amino acid possibilities at a single site.

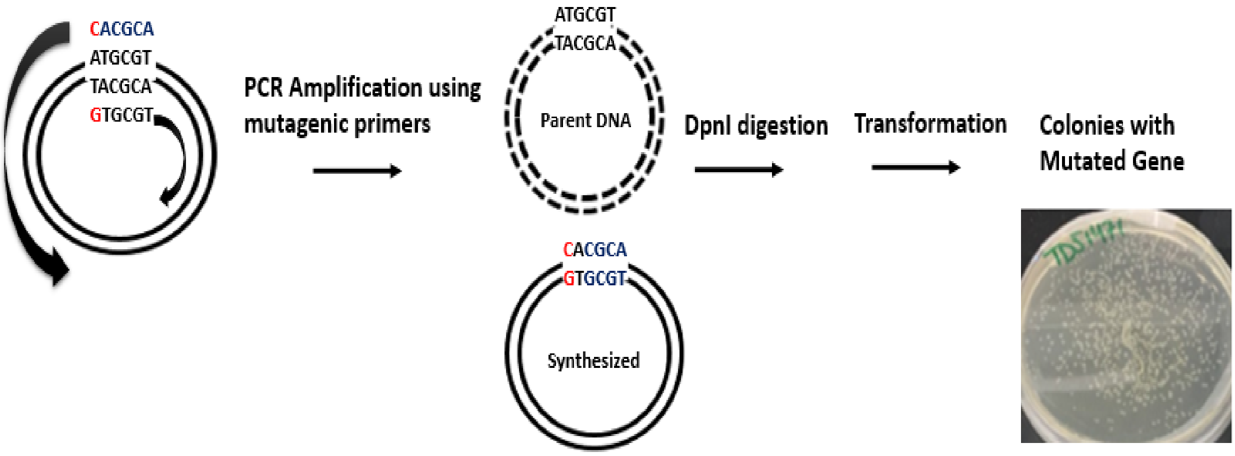

CACGCA
ATGCGT
TACGCA
GTGCGT
PCR Amplification using mutagenic primers
ATGCGT
TACGCA
Parent DNA
CACGCA
GTGCGT
Synthesized
DpnI digestion
Transformation
Colonies with Mutated Gene

Figure 1.5 Site directed mutatgenesis process. Primers containing the codon for a histidine CAC complements the parent plasmid containing codon for methionine ATG and is extended in PCR to generate a new strand of DNA. The newly synthesized DNA strand represents the new mutated DNA. Parent DNA is destroyed by DpnI digestion and the resulting mutated gene is transformed into E.*coli* to produce colonies with mutated gene.

1.2.4 Protein Expression and Purification Protocols

There are many advantages and challenges associated with protein expression and purification. To initiate expression, one must first answer the following questions: Which organism to use? Which plasmid is the most suitable? What is the most appropriate host? It is important to keep in mind that there are many options too choose from when answering these questions.

The idea of protein expression is that a gene of interest is cloned into a vector, transformed into bacteria, induced, expressed, and purified.[47-48] To achieve great success, identifying an organism is an essential first step. There are a variety of organisms commonly used as an expression system such as yeast and *E. coli*.[49] Both microorganisms have their advantages and disadvantages, however in this dissertation, *E. coli* bacterial cell lines are used to express all engineered proteins. *E. coli* bacterial cell lines offers high cell densities, fast growth kinetics, and simple transformation.[50]

Several factors must be considered when choosing a plasmid such as copy number and the incorporation of other genetic elements. Most expression vectors today contain an origin of replication, promoters, multiple cloning sites, selection markers, and affinity tags.[48, 51] The T7 expression system is the most popular approach for expressing proteins. In this system, a plasmid carrying the DNA encoding the protein cloned downstream the T7 promoter is introduced into an expression host (*E. coli* BL21 (DE3)). The host contain the T7 RNA polymerase gene which is regulated by a lac operon.[52] Thus, protein expression can be induced by molecules such as lactose or its non-hydrolyzable analog Isopropyl-β-D-1-thiogalactopyranoside (IPTG). IPTG remains constant throughout induction. Unlike IPTG, another induction method autoinduction can yield higher protein production.[53] Expression with ZYM media is induced with lactose. Cells must use up all the glucose in the media as the primary carbon source, then once depleted they will take up the lactose for expression.

There are challenges with protein expression such as protein degradation, protein misfolding, and poor solubility. Many proteins remain insoluble after induction causing them to form inclusion bodies. One can obtain soluble recombinant proteins by adding fusion partners such as SUMO[54] or adding molecular chaperones as supplementation.[55-57]

Post expression, one must be able to selectively isolate the protein of interest from a mixture of proteins through a process known as protein purification. Protein purification is crucial to understanding protein function.[58] There are several purification techniques that can be used such as gel filtration chromatography, ion-exchange chromatography, and affinity chromatography.[59] With these methods, proteins can be purified based on size and charge.

In this dissertation, all proteins were purified by nickel NTA affinity chromatography, others in chapter 3 by anion exchange chromatography using Q-Sepharose resin. The vector used has a small DNA portion encoding a poly His_6 tag that is usually located at the N-terminal or C-terminal of the protein sequence. This small peptide poly His_6 affinity tag is a feature of the vector construct that makes it extremely feasible for one step affinity purification.[60-61] The tag can bind with high affinity to nickel-NTA resin, while other proteins expressed in *E. coli* will flow straight through the column. The protein of interest is eluted from the column with high concentration of imidazole which will out compete the histidine tag for binding to the nickel-NTA resin. As a result of this, the protein will come off the column (Figure 1.6).

Generally the His-tag is followed by a coding sequence for tag removal such as a Tobacco Etched-virus (TEV) cleavage site.[62] Both of these sequences precedes the actual gene sequence. Therefore, TEV digestion would remove the His-tag and can result in increased protein purity and function. Enzyme kinetics can be done on a tagged protein or untagged protein and compared for catalytic rate enhancement as further discussed in Chapter 2 of this work.

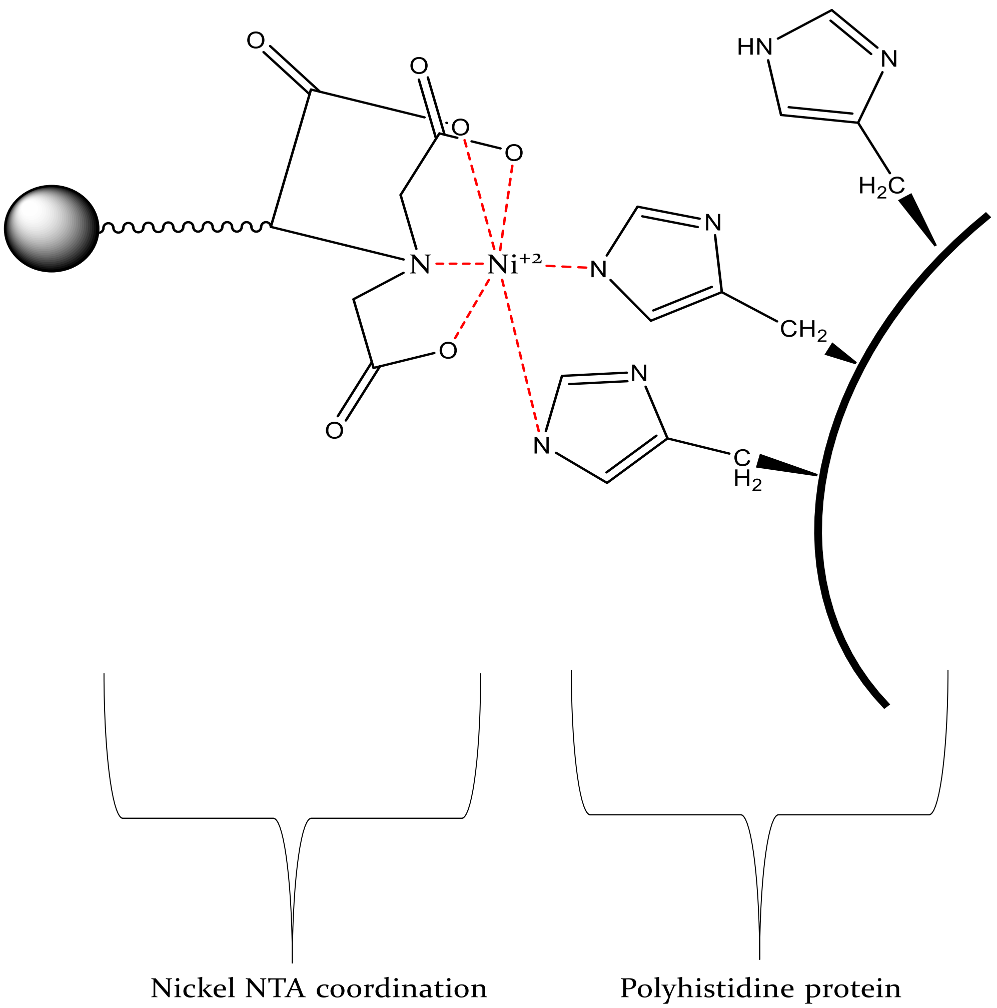

O
O
O
O
O
O
N
Ni+2
N
N
HN
N
H2C
CH2
N
N
C
H2
Nickel NTA coordination
Polyhistidine protein

Figure 1.6 Nickel NTA affinity chromatography coordination. NTA agarose tightly chelates nickel metal ion. The polyhistidine tag in the protein sequence coordinates to the nickel displaying high binding affinity. His tagged proteins bind to the column in this fashion while nonspecific proteins without the tag flow through.

1.2.5 Metalloprotein Design

As previously mentioned, the function of some enzymes are enhanced with the help of metal ions. They play both structural (i.e. zinc in zinc finger domains) and enzymatic roles (i.e. zinc in carbonic anhydrase). Key metal ions such as copper, zinc, and calcium, can serve as effector molecules that can bind to an allosteric site on the protein while also stabilizing it through a series of secondary interactions such as hydrogen bonding and electrostatic interactions.[13, 63]

Protein design strategies incorporate the fundamentals of protein structure and function for metalloprotein catalysis.[64] Metal sites in native proteins can efficiently catalyze many complex reactions in nature.[14] Although it can be challenging to modify metal sites for *de novo* enzyme design, there are several functional metalloproteins today with increased enzymatic activities. One can initiate the design approach in two ways: by starting with a known protein scaffold and introducing a novel metal site or engineering an existing site.[11, 14] With these opportunities for great catalytic success, metalloenzymes are suitable attractive targets for protein engineering.

1.3 Summary

In summary, enzymes are essential for regulating important chemical reactions to sustain life. Enzymes are often composed of amino acids that defines their primary structure and function. They function by lowering the activation energy of a system, thus stabilizing the transition state intermediate. Catalysis takes place in the protein's active site which involves catalytic residues and other secondary interactions from co-factors for active site stabilization.

Protein engineering is an attractive field for generating novel functions in existing proteins. Directed evolution is a technique used to generate mutations in a gene sequence through iterative cycles of diversification, screening, and amplification. Screening of combinatorial libraries and identification of target candidates has yielded catalysts with improved activities. Directed evolution approaches coupled with computational enzyme design has led to major advancements in *de novo* protein design. It aids in enhancing traditional computational strategies for predicting protein structure and function.

Protein design via a minimalistic approach simplifies complex computational design strategies by utilizing a computationally inexpensive approach to create new functionalities in proteins from their noncatalytic counterparts.[65] The goals of the projects described in this thesis are to employ a minimalistic approach for *de novo* designing efficient allosterically regulated catalysts for different chemical transformations and to understand the fundamental principles that govern enzymatic catalysis.

2. Directed Evolution of an evolvable catalyst for an unnatural reaction

The computational work from this chapter was done by the following authors and has led in part to the following publication.

Korendovych, I. V.; Kulp, D. W.; Wu, Y.; Cheng, H.; Roder, H.; DeGrado, W. F., Design of a switchable eliminase. *Proc Natl Acad Sci U S A* **2011,** *108* (17), 6823-6827.

2.1 Introduction

Over the past few years, protein engineering has gained recognition as a leading tool for understanding protein folding, structure, and function. Research has been done in investigating the principles of protein design strategies in addition to understanding the physiochemical properties of enzyme catalysis. Proteins can fold into well-defined structures and act as biocatalysts for reactions, but for some the catalytic rate remains low. Although a success, enzyme optimization still remains a challenge due to enzyme size, substrate range, and other external factors that can affect an enzyme's ability to function properly.[66] Computational enzyme design with directed evolution has become remarkably useful tool to overcome these limitations.[38, 67] There are several methodologies about creating new activities in proteins. Lots of approaches were tried, but many rely on extensive computational work done using Rosetta software.[30, 68]

However, it does suffer from several drawbacks (1) it relies mostly on geometric parameters obtained through quantum mechanics calculations for the transition state and (2) does not account for protein scaffold dynamics and size tuning of the active site in proteins.[69] The objective of this work is to focus on basic physicochemical principals such as substrate binding and specificity, functional group tuning, and scaffold stability to achieve catalysis using a minimalistic approach.[65]

2.2 De novo designing an artificial Kemp catalyst

2.2.1 Design Strategy

For the initial design approach, it is essential to select a suitable starting protein scaffold for evolution. There are many features that needs to be considered before selecting an appropriate target scaffold ranging from its size, to substrate binding abilities and thermodynamic stability.[35] Calmodulin (CaM) is a calcium binding regulatory protein that is ubiquitous in all eukaryotes.[70-72] It was selected as an initial framework for subsequent evolution due to fulfilling the requirements of an efficient catalyst. CaM is a versatile and promiscuous protein due to its flexibility in binding a variety of different substrates and target proteins in its hydrophobic cavity. That flexibility is due to having a nonpolar featureless hydrophobic microenvironment and the flexible methionine residues are responsible for accommodating a variety of different molecules. Calmodulin is composed of two α- helical binding domains connected by a flexible linker. Each helical domain contains two EF hands that can bind two calcium ions. When calcium is not bound, CaM is in a closed state, but when calcium is bound, CaM undergoes a conformational change (Figure 2.1).[73] Because of this binding, calmodulin's fold is very stable. The fundamental *question is: Can* a protein be effectively converted into a catalytically active enzyme using only the minimal amount of mutations? To answer this question, employed a simple minimalistic approach was undertaken to test whether CaM can adopt catalytic activity.

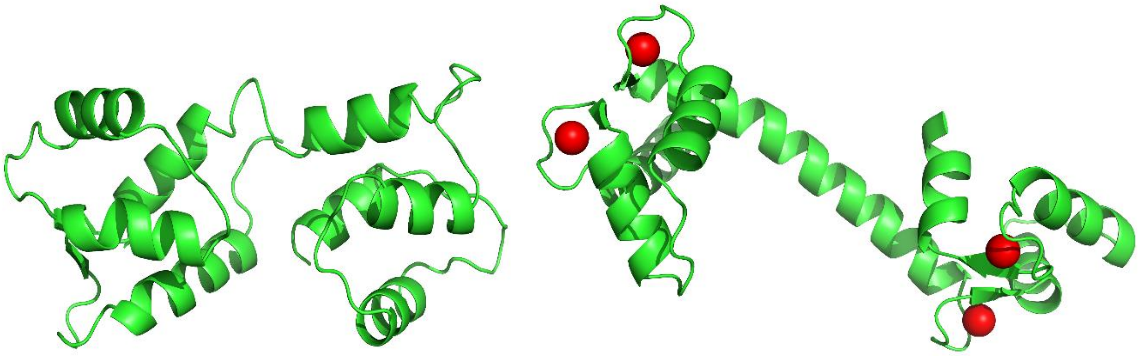


Figure 2.1 Unbound state of calmodulin (left) PDB 1CFC. Bound state of calmodulin with calcium ions as red spheres, PDB 3CLN.

Many reactions and chemical transformations exists in nature and are catalyzed by a variety of different enzymes. Our reaction of interest was Kemp elimination reaction.[70-71, 74] This is an unnatural reaction that is commonly employed as a model reaction for enzyme catalysis. It involves a C-H bond abstraction by a catalytic base to generate a single high energy transition state (Figure 2.2). The product of the reaction can be spectroscopically measured at 380 nm.[75-76]

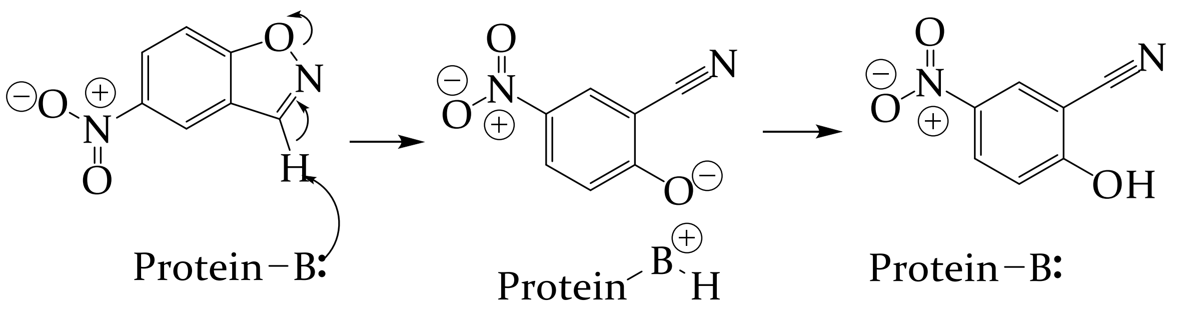


Figure 2.2 Kemp elimination reaction. 5-Nitrobenzisoxazole as the substrate.

Dehydrated carboxylates are great basic catalysts for Kemp elimination reaction due to its basicity towards carbon acids. Therefore, an effective Kemp eliminase can be engineered by placing a single carboxylate at the bottom of the hydrophobic cavity to increase deprotonation.[70] Active site residues in the pocket were computationally identified to accomdate a single catalytic carboxylate residue (Figure 2.3). After the identification of the specified amino acids, Van der Waals interaction energies were computed to thermodynamically score the mutations for tolerating an aspartate or glutmate residue. The higher the score the less thermodyanically stable that mutation is in the protein. Although positions 105 and 125 demonstrated low van der Waals energies, position 92 was the most tolerable lowest energy rotamer sequence (Figure 2.4).

SLMKDTDSEE EIREAFRVFD KDGNGYISAA

ELRHVMTNLG EKLTDEEVDE MIREADIDGD

GQVNYEEFVQ MMTAK

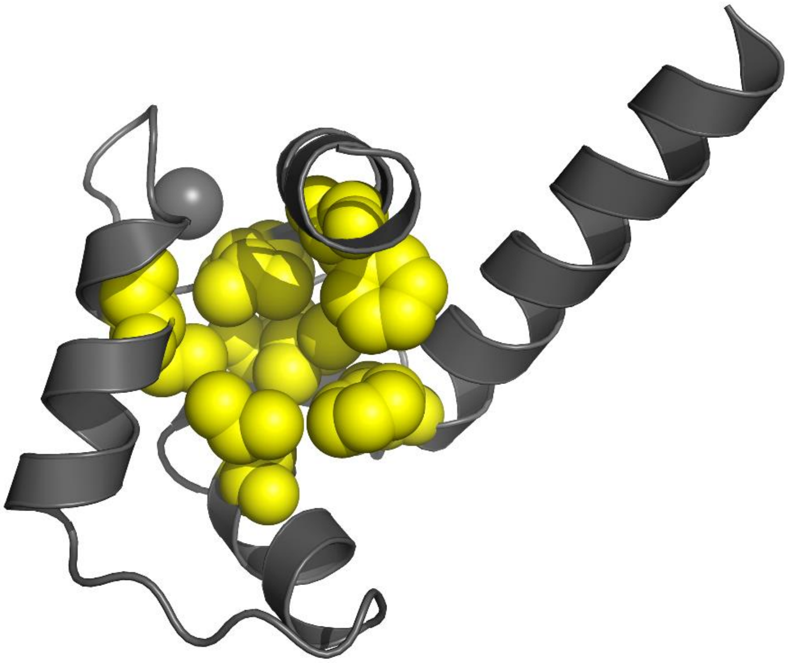


Figure 2.3 C-terminal domain of CaM with potential active site catalytic residues (yellow spheres) for accomodating carboxylates.

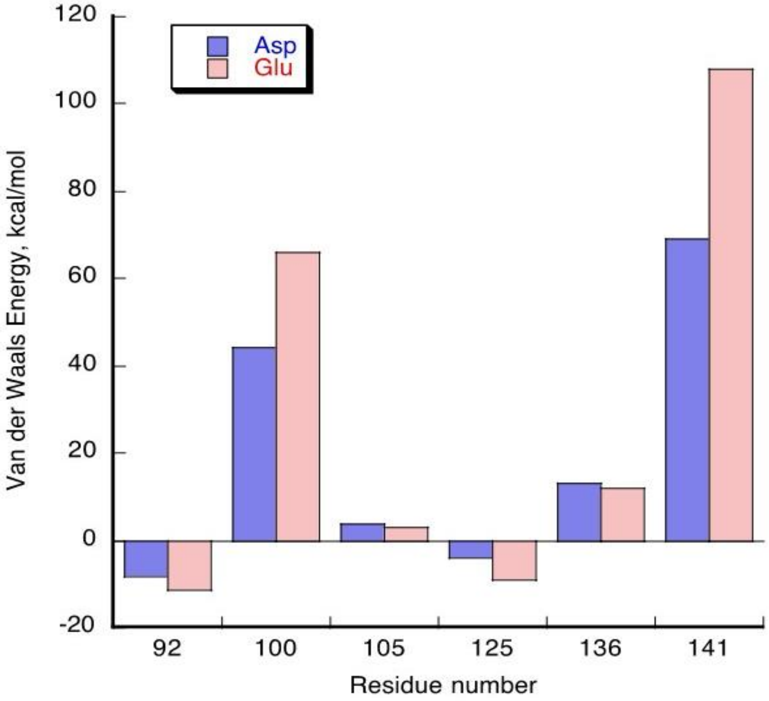


Figure 2.4 Van der Waals energies of Asp (blue) or Glu (pink) as lowest rotamers at each hydrophobic residue.

Next, the glutamate residues at all the identified positions for Michaelis active site were evaluated for stabilization of the transition state geometry. A superrotamer library was created where the carboxylate was virtually fused to the substrate in the transition state geometry and tested at each position.[70] Position 92 was the only one that could accomodate Glu in the calcium bound state in such a way that transition state geometry could be preserved (Figure 2.5). [70]

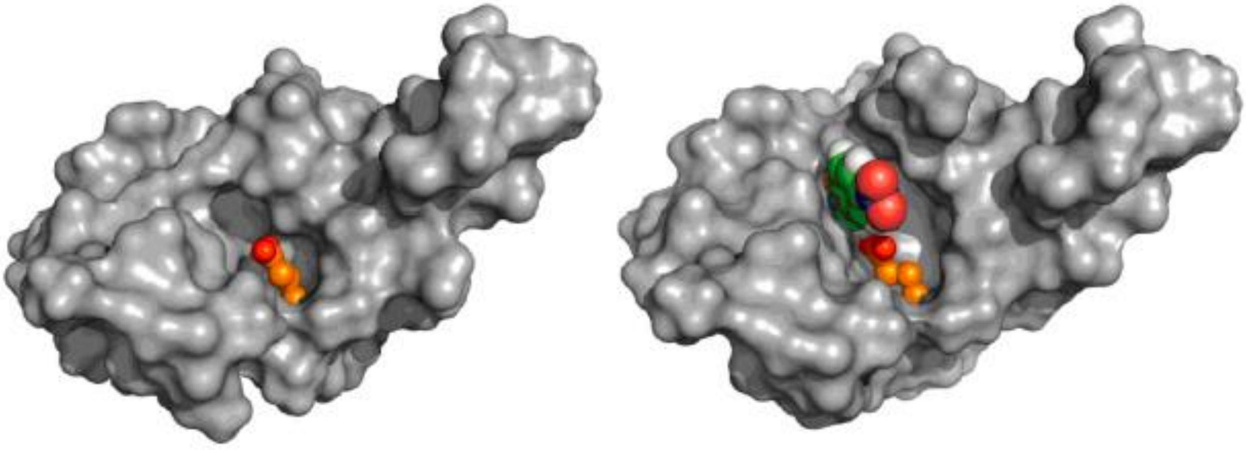


Figure 2.5 Left: Examination of the Glu residues in the identified positions/ docking of the substrate. Right: Transition state docking based on a search of a superrotamer library for the transition state.

Position F92 was mutated to glutamic acid and that mutation was characterized for catalyzing Kemp elimination. Both full length and c-terminal CaM have been kinetically characterized for Kemp elimination. Subjecting CaM F92E to several rounds of directed evolution produced improved activity. Each subsequent round of evolution after F92E was termed F92E round 1 to round 7 (these mutants are classified by R for Round number and are full length proteins). After 7 rounds of mutagenesis, a 200- fold increase in catalytic activity[71] generated a new variant AlleyCat7 (all AlleyCats are C-terminal proteins). See table 2.1 in experimental for efficiencies of evolved proteins.

2.2.2 Michaelis-Menten Kinetics

The Michaelis-Menten model,[77] commonly used to describe enzymatic catalysis is shown in equation 1: enzyme (E) reacts with substrate (S) to yield an intermediate (ES). The (ES) complex is converted to product which will then dissociates from the active site to form free enzyme (E) and (P).

$$[E] + [S] \rightleftharpoons [ES] \rightharpoonup [E] + [P] \qquad (1)$$

To study the behavior of enzymes and how they catalyze reactions, we must study the rates at which they catalyze the reactions. One way to study reaction rates is by the following graph below (Figure 2.6). The graph below displays V_o as the rate at which an enzyme catalyzes its specific substrate, and [S] as the concentration of the substrate that binds to the active site of the enzyme. Initially, when substrate concentration is low, the rate of the enzyme will be proportional to the concentration of substrate until it reaches V_{max} and level off. This type of enzyme behavior can be fit to the mathematical expression known as the Michaelis-Menten equation:

$$V_o = \frac{k_{cat}\,[E]\,[S]}{K_M + [S]} \qquad (2)$$

According to this equation, we made the steady state approximation that [ES] is constant throughout the reaction. k_{cat} (turnover number) provides information about how many substrate molecules is converted to product per unit time, K_M is the substrate concentration that allows the enzyme to achieve half its maximal velocity, and k_{cat}/K_M is the enzyme's catalytic efficiency.[77] All data in this dissertation was fit to the linear regime as shown in equation 3 of the Michaelis- Menten equation unless otherwise noted.

$$V_o = \frac{\mathrm{k}_{\mathrm{cat}}\,[E]\,[S]}{\mathrm{K}_{\mathrm{M}}} \qquad (3)$$

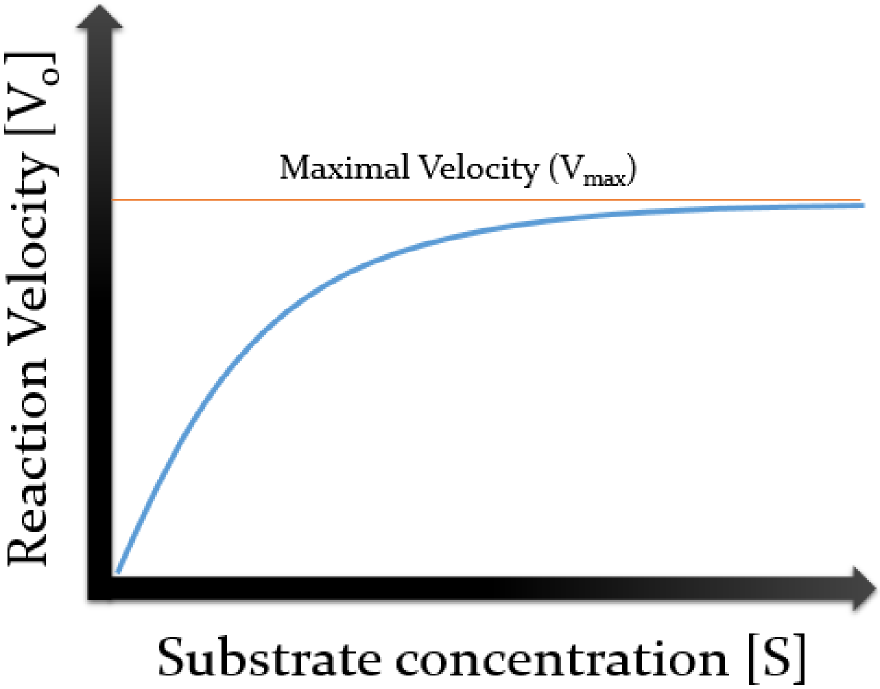


Figure 2.6 Enzyme's behavior demonstrated by the Michealis Menten curve. Reaction velocity as a function of substrate concentration.

2.2.3 Directed Evolution of CaM F92E 7th R for Enhanced Activity

Directed evolution methods has aided in improving a scaffold's initial catalytic activity. In our directed evolution experiments, we are interested in identifying the activity limit for calmodulin full length and C-terminal proteins. We work with both forms of the protein for different reasons. The full lengths proteins provide the same catalytic efficiencies as their C-terminal counterparts, but offer higher expression yields that aids in characterization.

For furthering evolution, our starting scaffold of interest was CaM F92E 7th R. Previously we have characterized this protein and obtained a catalytic efficiency of~800 $M^{-1}s^{-1}$ at pH 7.[71] Our goal was to subject this protein to more rounds of evolution without destabilizing the protein's structure. Error Prone PCR (ep-PCR) uses a polymerase with a high error rate to introduce random mutations throughout a gene of interest.[78] This method was employed for analyzing the effects of random mutations throughout CaM F92E 7th R on its activity for Kemp elimination (see experimental for details).

From the ep-PCR, several hits were sequenced and nine mutants were identified as D131G, N111D, V142L, D133N, T110R, G96C, D80H, K94M, and S81R. All mutants were characterized for Kemp elimination over the substrate range of (0.00012-0.00096 M), and the best variant identified on par with wild-type CaM F92E 7th R was S81R (Figure 2.7). The second best was K94M (Figure 2.8). This showed us that a positive charged residue arginine (R) and hydrophobic side chain methionine (M) in the C-terminal domain of CaM can promote catalytic activity. Since S81R was the best mutant, it was recloned into pMCSG49 vector and the catalytic activity was re-evaluated for only the c-terminal fragment and compared to its full length construct (Figure 2.7). Graphs for D131G, N111D, V142L, D133N, T110R, G96C are shown in appendix at the end of this chapter. Please note that Kemp substrate is not soluble past 1 mM.

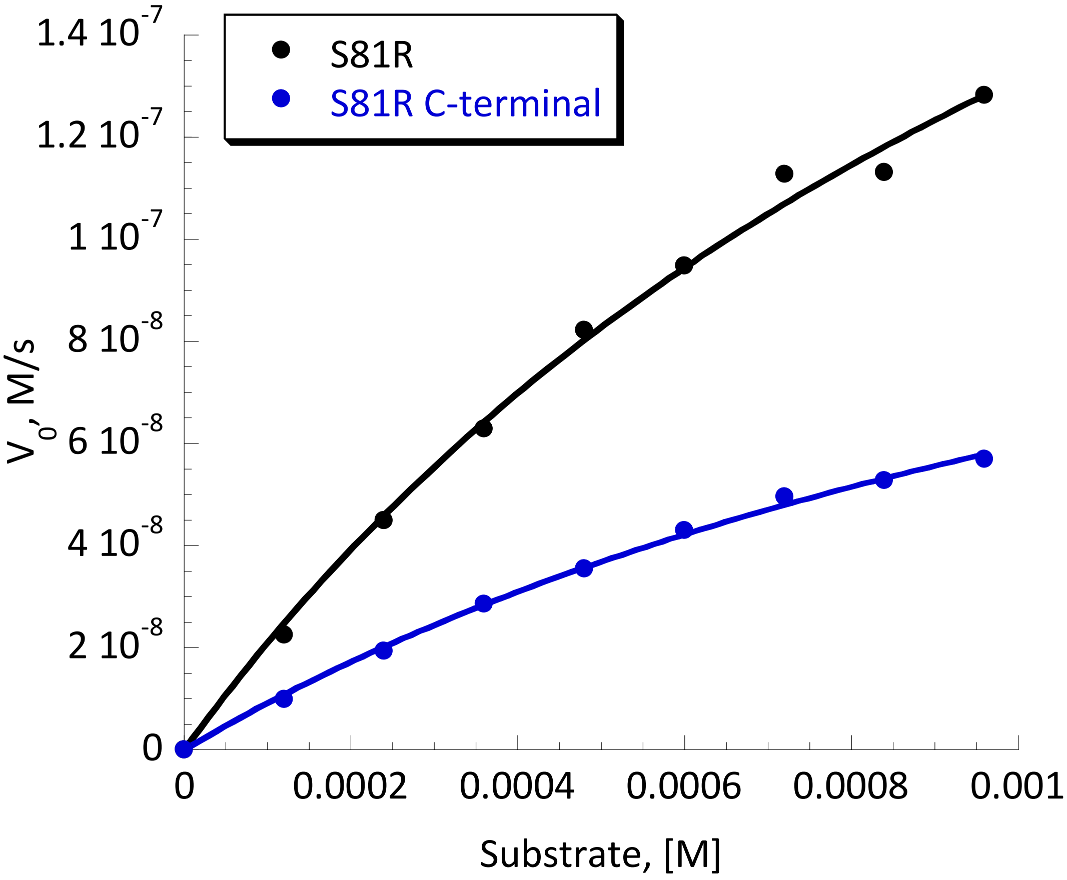


Figure 2.7 Kemp elimination activities of S81R CaM F92E 7th R and of S81R CaM F92E 7th R C-terminal mutant with affinity tag. Conditions: 20 mM HEPES (pH 7.0); 100 mM NaCl; 10 mM $CaCl_2$; substrate concentration 0.00012-0.00096 M; 0.2 μM S81R CaM F92E 7th R; 0.18 μM S81R CaM F92E 7th R C-terminal; the catalytic efficiencies obtained from fitting the linear portion of the curves are k_{cat}/K_M 882 ± 1.6 $M^{-1}s^{-1}$, k_{cat}/K_M 447 ± 6.5 $M^{-1}s^{-1}$ respectively. (Unpublished, TD1054 and TD1103)

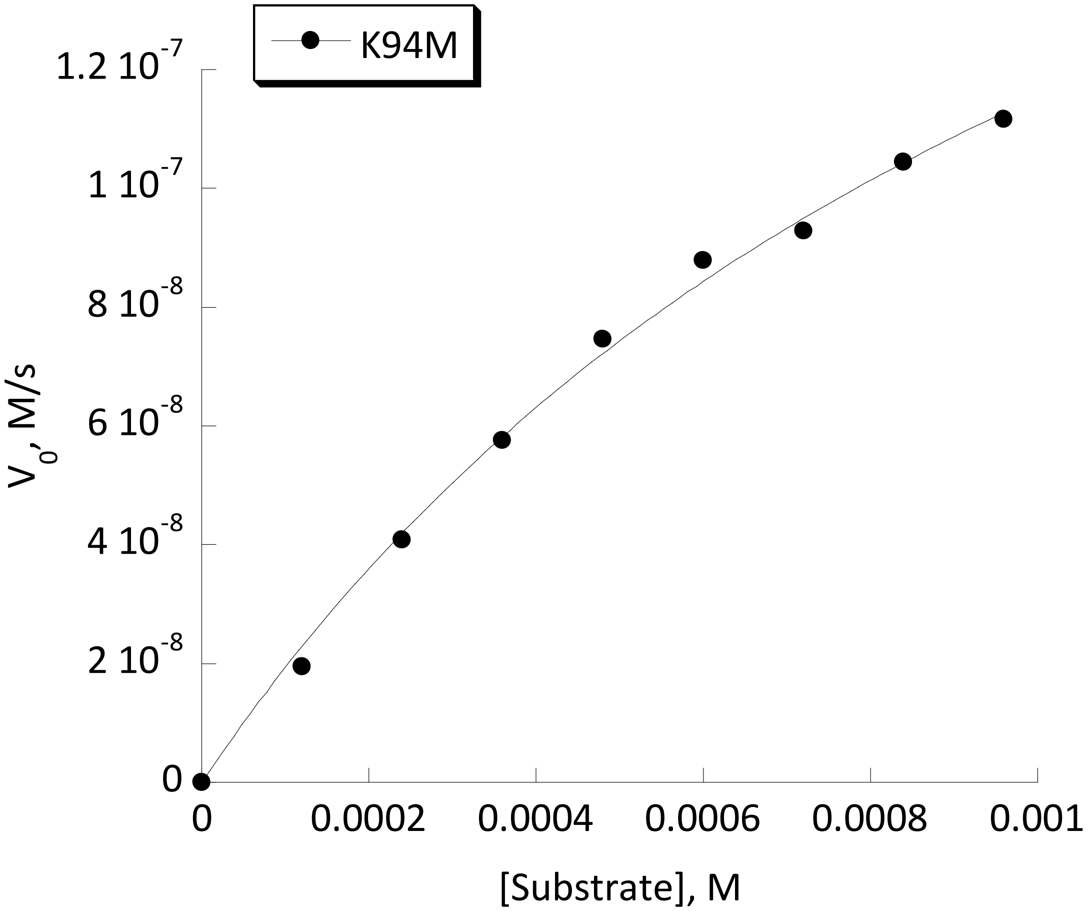


Figure 2.8 Kemp Elimination activity of K94M CaM F92E 7th R mutant with affinity tag. Conditions: 20 mM HEPES (pH 7.0); 100 mM NaCl; 10 mM $CaCl_2$; substrate concentration 0.12-0.96 mM; 0.2 μM K94M CaM F92E 7th R ; the catalytic efficiencies obtained from fitting the linear portion of the curves are (k_{cat}/K_M = 834 ± 20 $M^{-1}s^{-1}$). (Unpublished, TD10621)

Often times the His tag on proteins can influence how well an enzyme can function. It is important to remove the tag and compare catalytic efficiencies to those with the purification tag. For the two best mutants mentioned above, the tags were removed during purification and the activities were evaluated in a single run (Figure 2.9). Removing the His-tag in these cases did not augment the catalytic activities compared to their activities with the affinity tag.

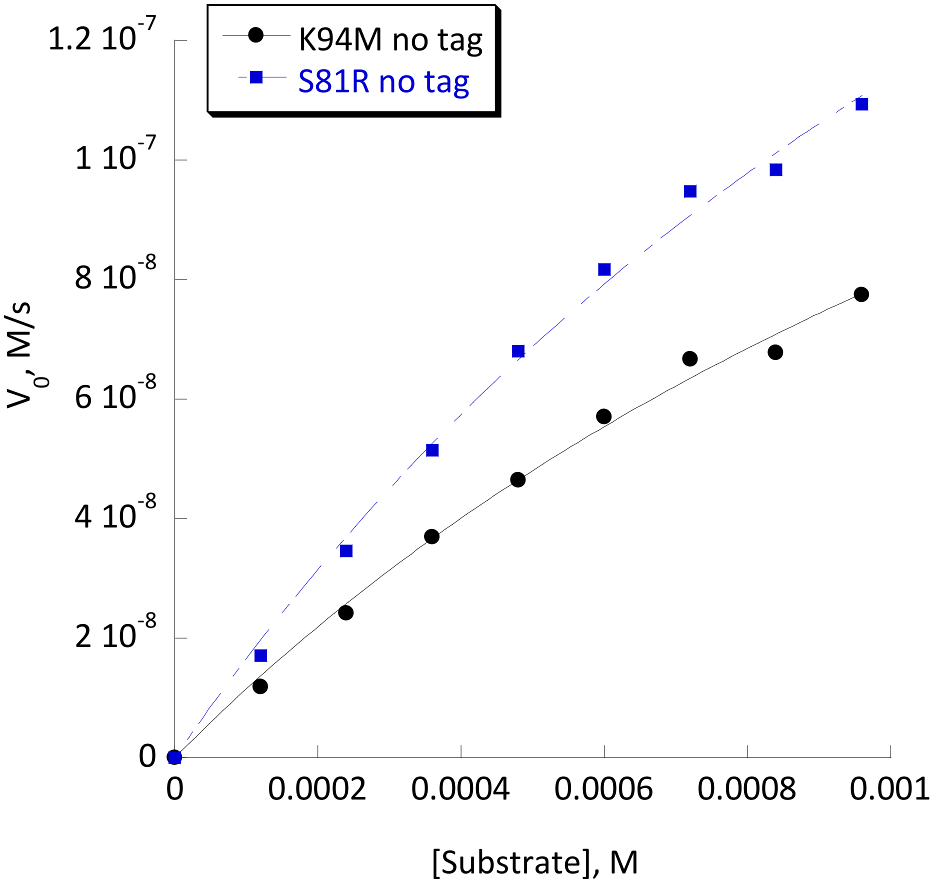


Figure 2.9 Kemp Elimination activity of S81R & K94M CaM F92E 7th R mutant without affinity tag. Conditions: 20 mM HEPES (pH 7.0); 100 mM NaCl; 10 mM CaCl2; the catalytic efficiencies obtained from fitting the linear portion of the curves are $k_{cat}/K_M = 721 \pm 5.3\ M^{-1}s^{-1}$ and $504 \pm 5\ M^{-1}s^{-1}$ respectively. (Unpublished, TD10733)

2.2.4 pK_a of the catalytic residue

The pK_a is related to amino acid side chains specifically their ionization properties. Since enzymes activity and stability is pH dependent, it is important to analyze the changes in pK_a of amino acid side chains. These changes occur when amino acids in the active site of the protein form interactions with other neighboring residues. Therefore, we are interested in examining the pK_a shifts (increasing or decreasing) of the glutamate residue as we introduce new mutations post 7th R. The $(k_{cat}/K_M)_{max}$ values were obtained from a fit of pH-profiles.

CaM F92E has a pK_a of 6.98 ± 0.05. Active glutamate residue in the evolved CaM F92E 7th R decreased substantially to 6.43 ± 0.02 (Figure 2.10).

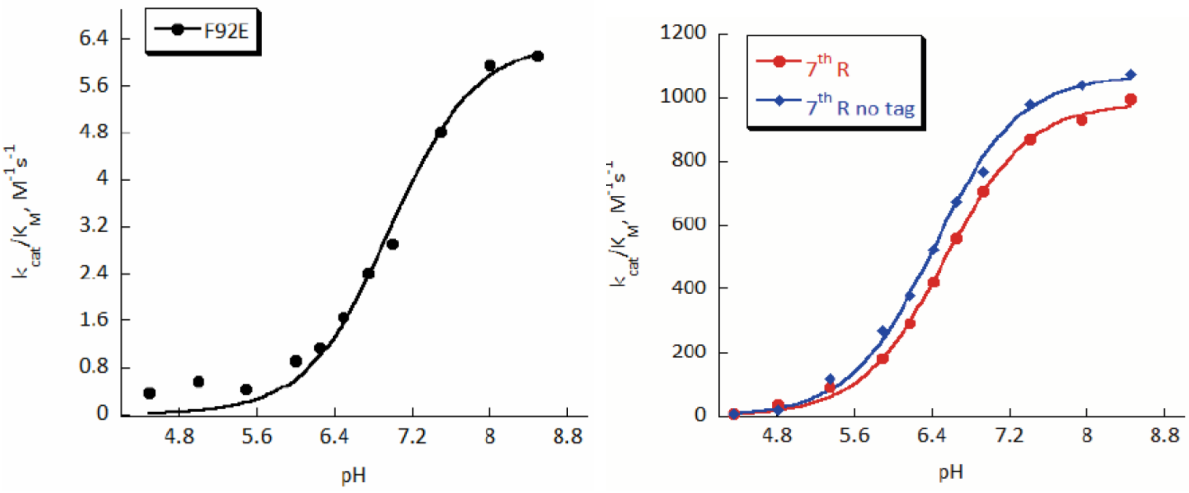


Figure 2.10 pH profiles of original scaffold CaM F92E (left) and CaM F92E 7thR (right). Conditions: pH 4.5-8.5, 150 µM substrate.

The activity of previously mentioned mutants obtained from the ep-PCR screening described above, was studied as a function of pH. Variant S81R displayed the highest $(k_{cat}/K_M)_{max}$ value of 1285 ± 8 compared to other variants and slightly higher the original scaffold CaM F92E 7th R which had a $(k_{cat}/K_M)_{max}$ of 1069±14. The pK_a of the glutamate residue in CaM F92E 7th R did not change with the introduction of arginine (6.4 ± .01) Moreover, variant S81R has the same $(k_{cat}/K_M)_{max}$ and pK_a as AlleyCat 7 which was 1283 ± 13 and 6.39 ± 0.02 respectively (Figure 2.11). With affinity tag removed, S81R displayed a lower $(k_{cat}/K_M)_{max}$ of 1192±23 and a very small difference in pK_a compared to S81R with affinity tag.

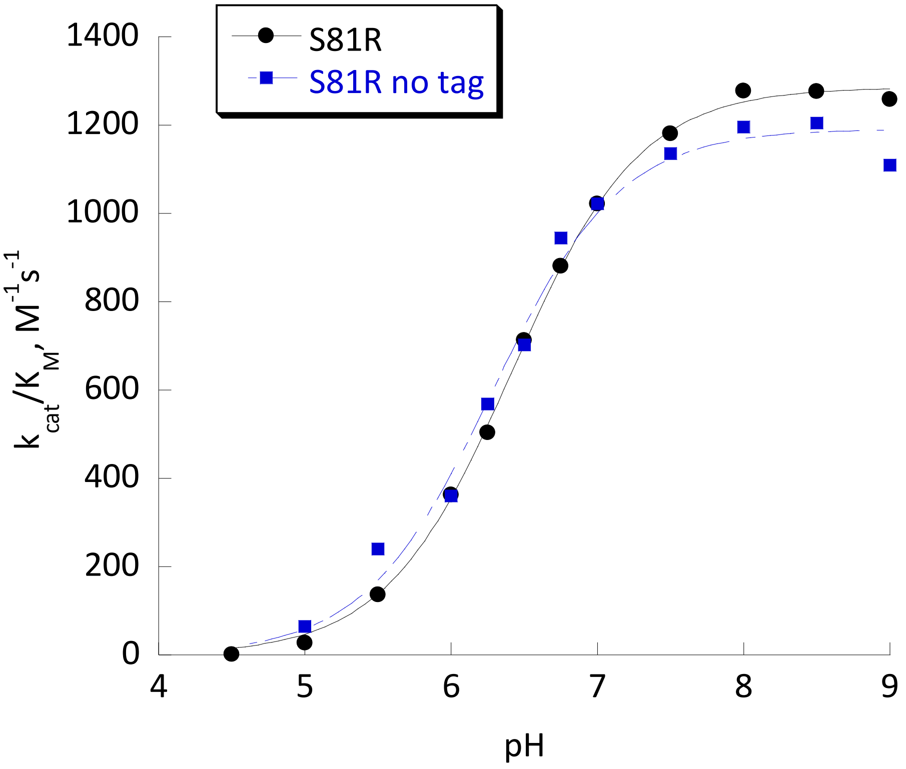


Figure 2.11 pH profile of S81R with and without affinity tag. Conditions: pH 4.5-9.0; 150 µM substrate; 0.20 µM S81R. (Unpublished, TD10583 and TD10776)

For the second best variant K94M, has a $(k_{cat}/K_M)_{max}$ value of 1200±24 but there was a small increase in the pK_a 6.5±0.03. With affinity tag removed, K94M displayed a lower $(k_{cat}/K_M)_{max}$ of 976±21 and a decrease in pK_a (6.38±0.04) compared to K94M with affinity tag (Figure 2.12).

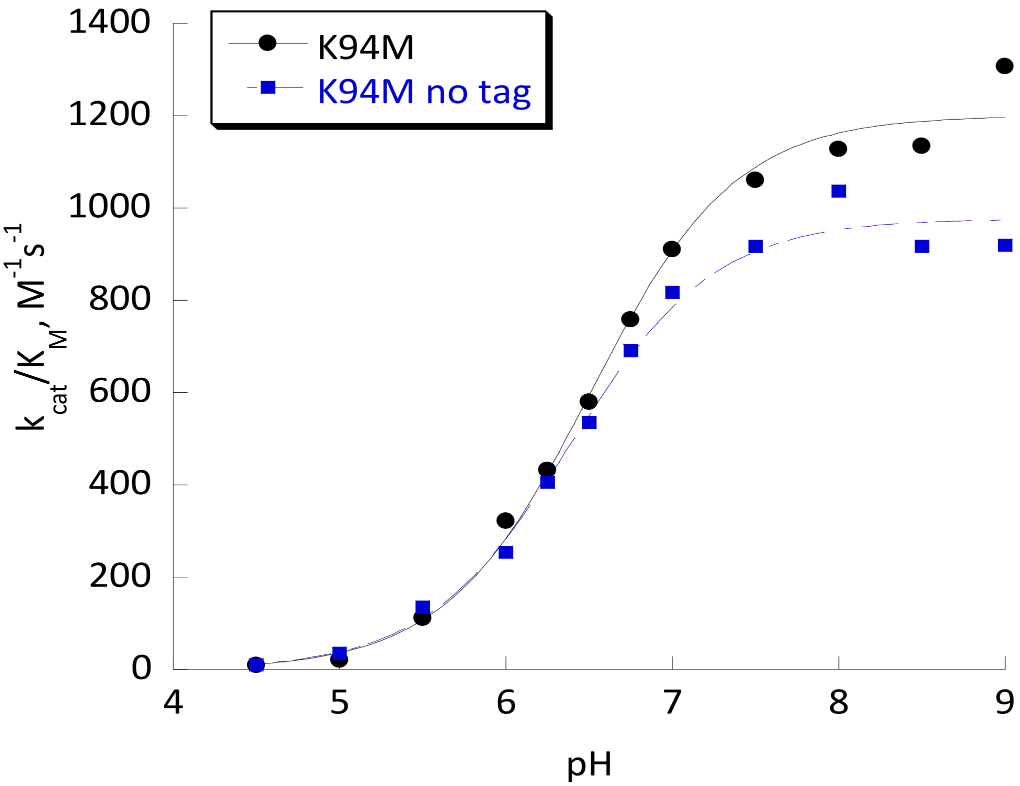


Figure 2.12 pH profile of K94M with and without affinity tag. Conditions: pH 4.5-9.0; 150 μM substrate; 0.20 μM K94M. (Unpublished, TD10591 and TD10894)

The introduction of leucine at position 142 resulted in a $(k_{cat}/K_M)_{max}$ of 1076 ± 10 and a small decrease in pK_a 6.3 ± 0.02 (Figure 2.13).

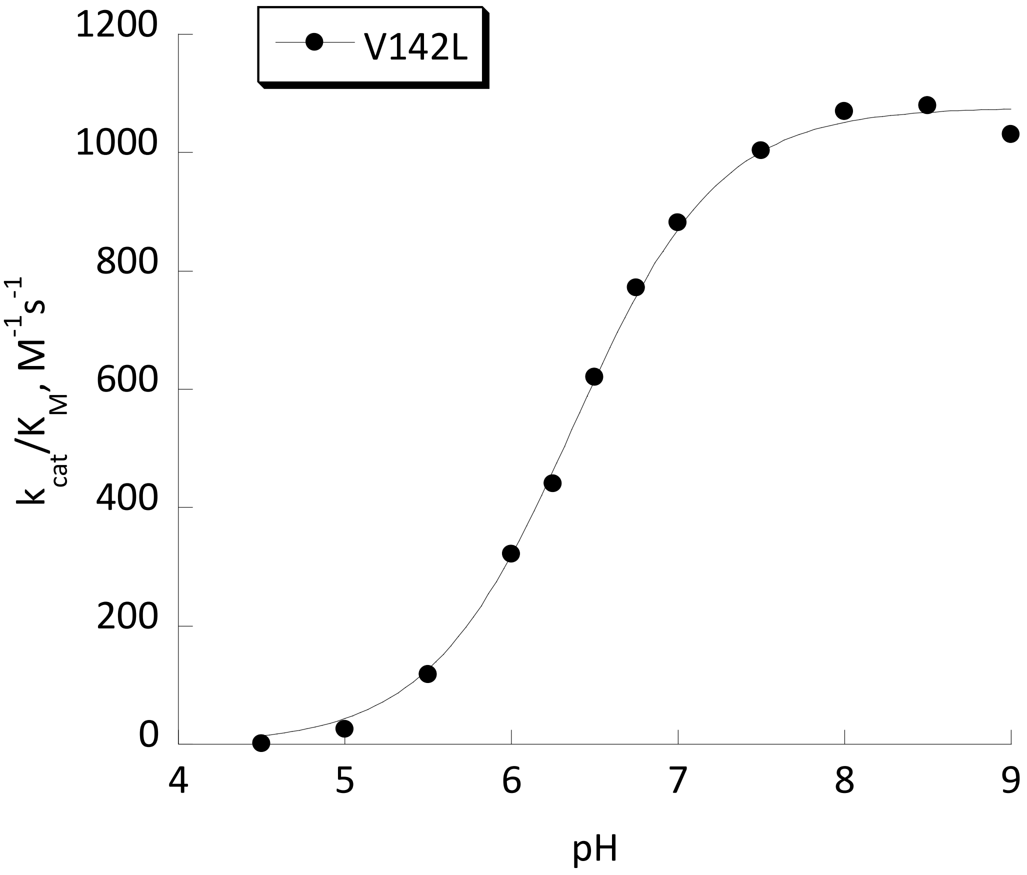


Figure 2.13 pH profile of V142L with affinity tag. Conditions: pH 4.5-9.0; 150 μM substrate; 0.20 μM K94M. (Unpublished, TD10583)

Variant D133N portrays very close $(k_{cat}/K_M)_{max}$ of 1061 ± 9 and pK_a 6.47 ± 0.01 to that of CaM F92E 7th R (Figure 2.14). Introducing an arginine at position 110 resulted in a $(k_{cat}/K_M)_{max}$ of 1134±8 and a very similar pK_a of 6.4±0.01 (Figure 2.15). The pK_a for S81R and T110R are identical.

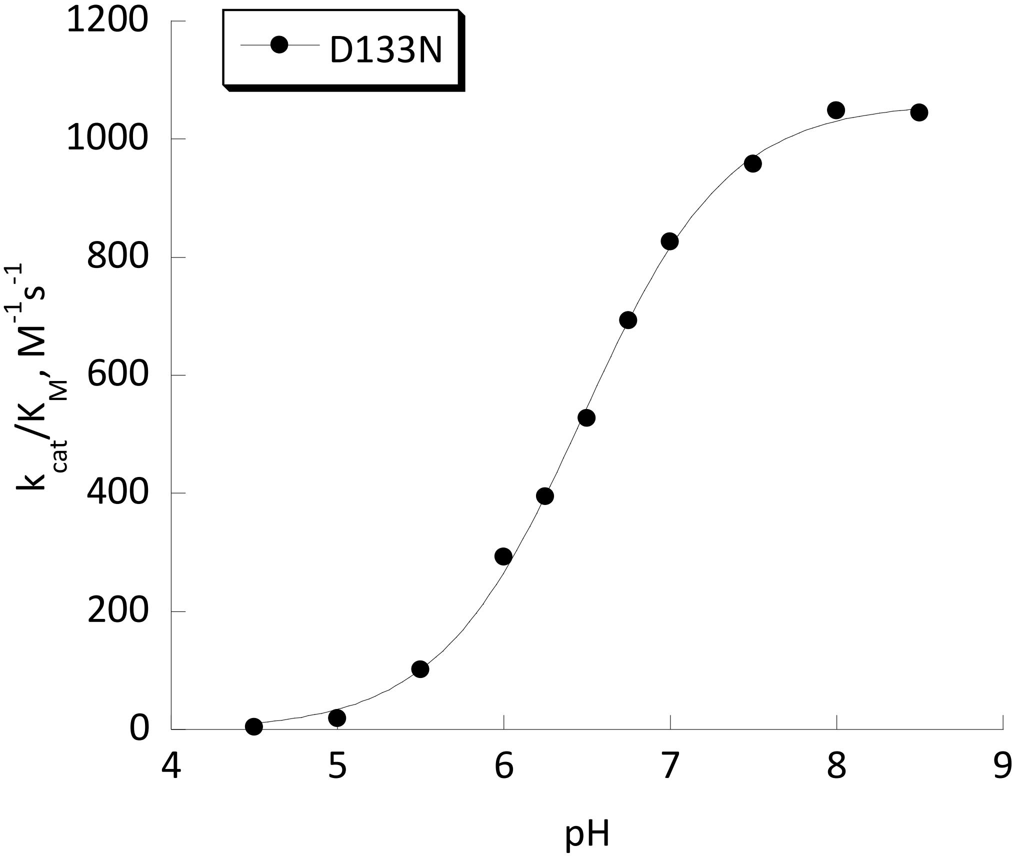


Figure 2.14 pH profile of D133N with affinity tag. Conditions: pH 4.5-9.0; 150 μM substrate; 0.20 μM D133. (Unpublished, TD10591)

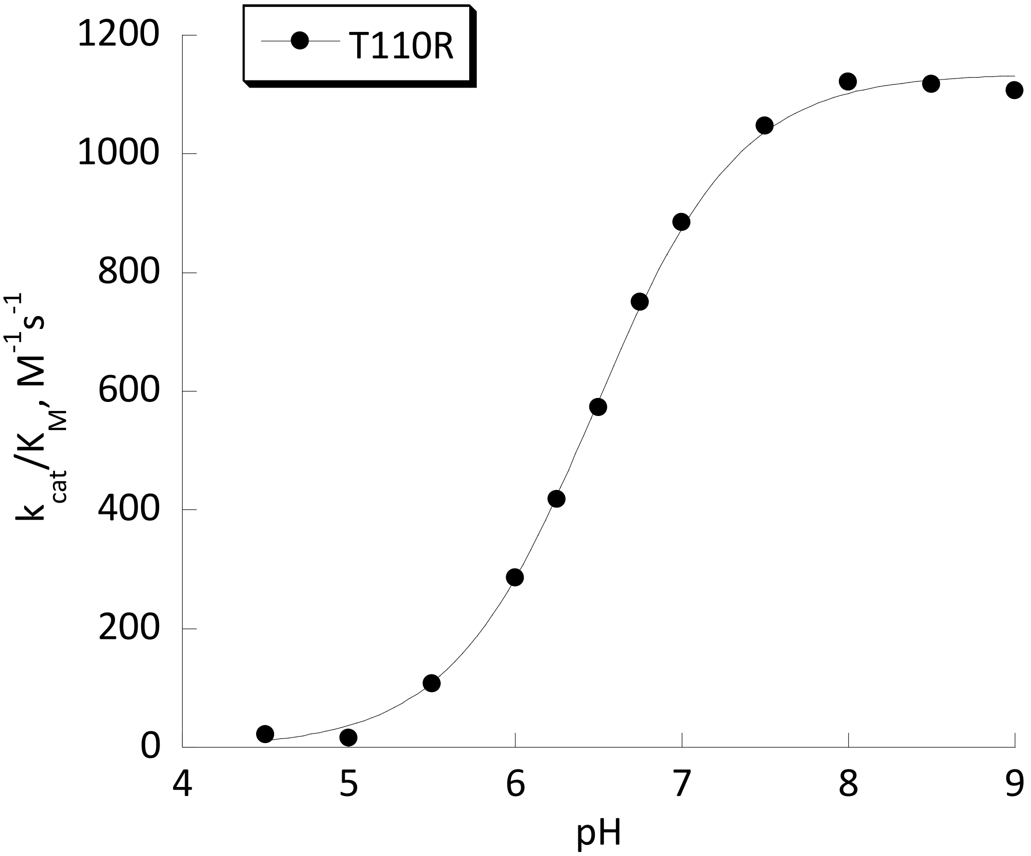


Figure 2.15 pH profile of T110R with affinity tag. Conditions: pH 4.5-9.0; 150 μM substrate; 0.20 μM T110R. (Unpublished, TD10592)

Table 2.2: Kinetic Data summary of ep-PCR variants evolved from CaM F92E 7th R.

7thR mutations	**k_{cat}/K_M $M^{-1}s^{-1}$**	**$(k_{cat}/K_M)_{max}$ $M^{-1}s^{-1}$**	**k_{cat} s^{-1}**	**K_M mM**	**pK_a**
No mutation	812 ± 8	1069 ± 14	1.86 ± 0.10	2.1 ± 0.2	6.43 ± 0.02
S81R	882 ± 1.6	1285 ± 8	1.5 ± 0.17	1.4 ± 0.24	6.4 ± 0.01
K94M	834 ± 20	1199 ± 24	1.2 ± 0.10	1.2 ± 0.16	6.5 ± 0.03
D133N	776 ± 16	1061 ± 9	1.2 ± 0.11	1.3 ± 0.18	6.47 ± 0.01
V142L	762 ± 12	1076 ± 10	1.4 ± 0.19	1.6 ± 0.31	6.3 ± 0.02
T110R	717 ± 26	1134 ± 8	1.1 ± 0.11	1.2 ± 0.18	6.4 ± 0.01
D131G	683 ± 0.75	-	0.63 ± 0.09	3.0 ± 0.55	-
N111D	590 ± 0.64	-	1.5 ± 0.12	2.3 ± 0.24	-
G96C	512 ± 27	-	0.99 ± 0.10	1.6 ± 0.24	-
D80H	504 ± 12	-	1.2±.13	2.0 ± 0.30	-

2.2.5 Calcium Dependence for Enzymatic Activation

Calmodulin is regulated in the presence of calcium, therefore to be certain that directed evolution did not destroy its affinity for calcium, a calcium dependence assay was completed for K94M and D133N only just to observe the metal concentration that induces the maximum catalytic activity. In both cases for K94M and D133N, showed that at least 0.2 mM calcium is required for approximately 0.2 μM enzyme to achieve maximum activity (Figures 2.16- 2.17).

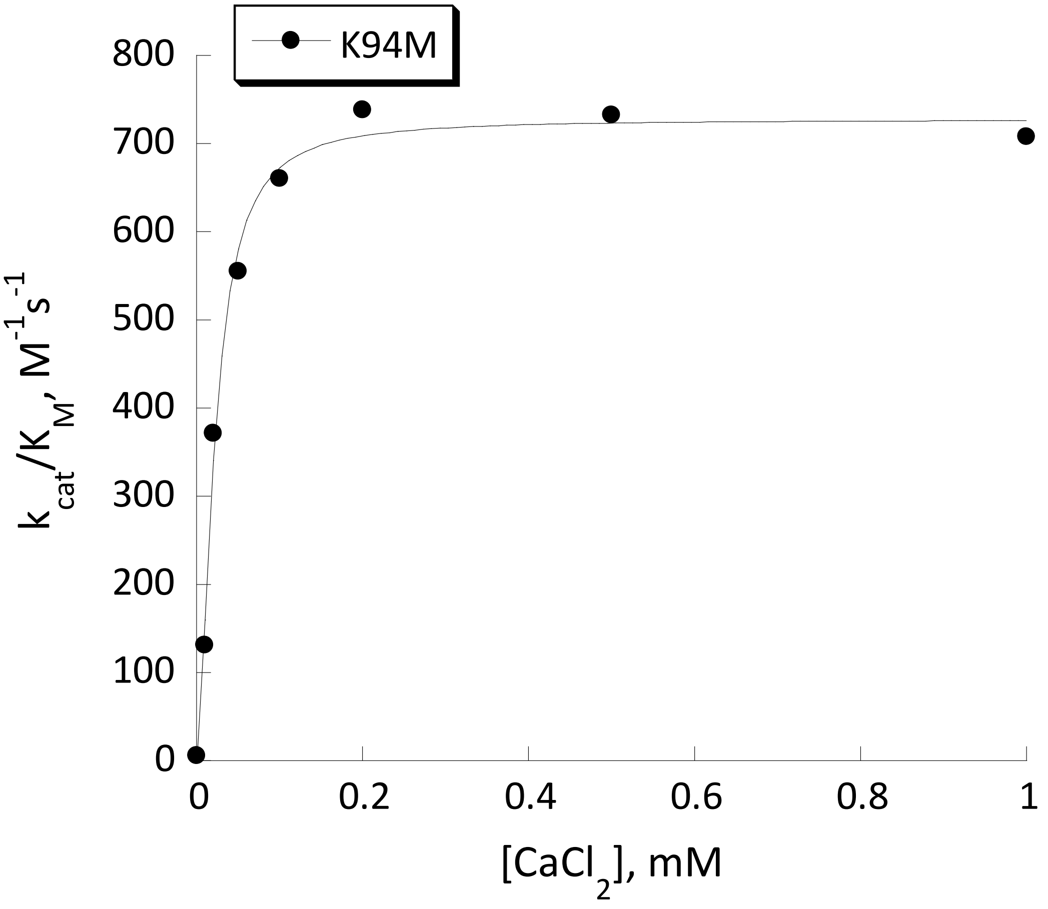


Figure 2.16 Dependence of the activity of K94M on [Ca^{+2}] at pH 7.0 in 20 mM HEPES; 100 mM NaCl; substrate concentration 100 μM; protein concentration 0.20 μM. $(k_{cat}/K_M)_{max}$ = 728 ± 18 $M^{-1}s^{-1}$, M_{50} = 21.8 ± 1.7 μM, h = 1.6 ± 0.20. (Unpublished, TD10681).

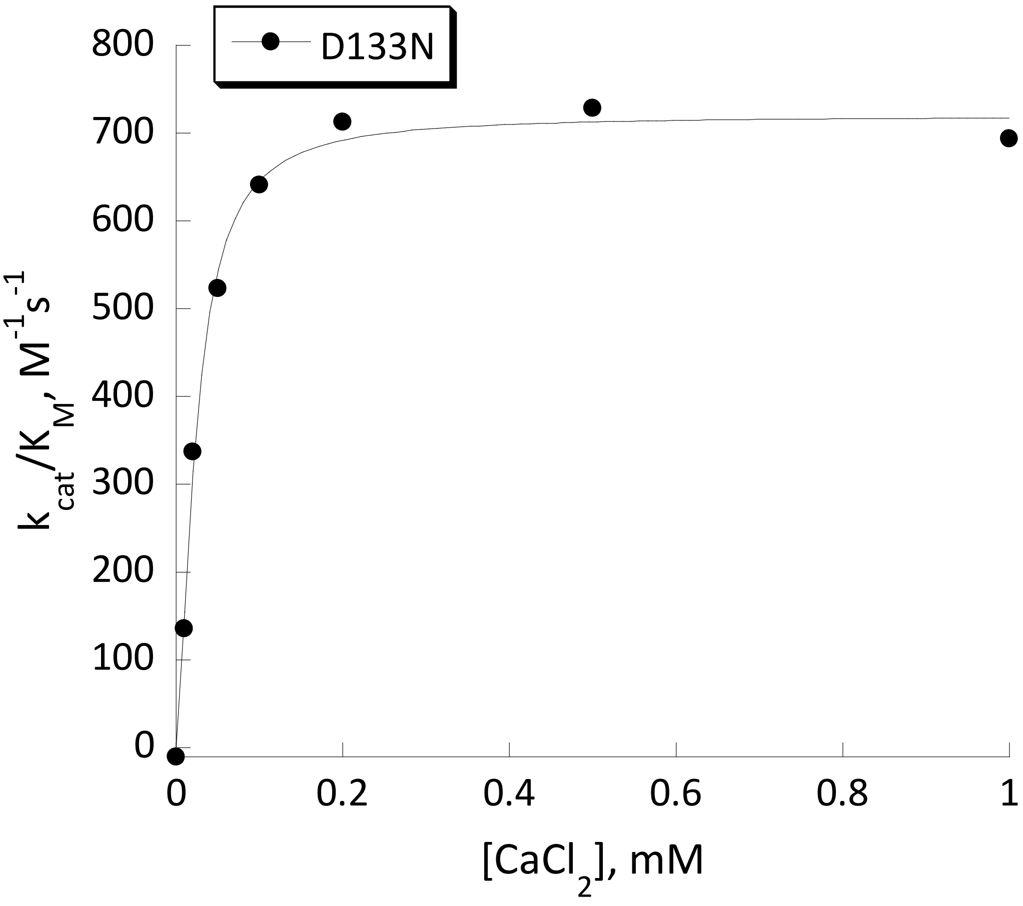


Figure 2.17 Dependence of the activity of D133N on $[Ca^{+2}]$ at pH 7.0 in 20 mM HEPES; 100 mM NaCl; substrate concentration 100 μM; protein concentration 0.20 μM. $(k_{cat}/K_M)_{max}$ = 720 ± 16 $M^{-1}s^{-1}$, M_{50} = 23.8 ± 1.7 μM, h = 1.5 ± 0.15. (Unpublished, TD10681).

Positions from saturation mutagenesis of CaM F92E 7th R with NNK codon degeneracy: Q88, V91, L105, I107, V108, R112, T128, R144. One variant obtained at position 88, Q88R (was not characterized). I107V was obtained and characterized below with a 2 fold lower k_{cat}/K_M 410± compared to 7th R (Appendix Figure S.09).

2.3 Conclusions

In conclusion, we have employed a minimalistic design approach with the aid of directed evolution to create an efficient catalyst of Kemp elimination. This computational enzyme design strategy involves selecting a non-enzymatic starting scaffold, analyzing the physical and chemical properties of the protein's active site, and finally computationally examining the interaction between the catalytic groups and the proposed reaction of interest. Calmodulin scaffold was tuned to increase its active site functionality, stability, and catalytic activity. Random positions throughout full length calmodulin's scaffold were explored to further improve catalytic efficiency in Kemp elimination. The mutants mentioned in this chapter did not perturb or inactivate calmodulin.

Appendix Chapter 2 Supplemental Experiments

Graphs for D131G, N111D, V142L, D133N, T110R, G96C shown below at pH 7.0 and 8.0.

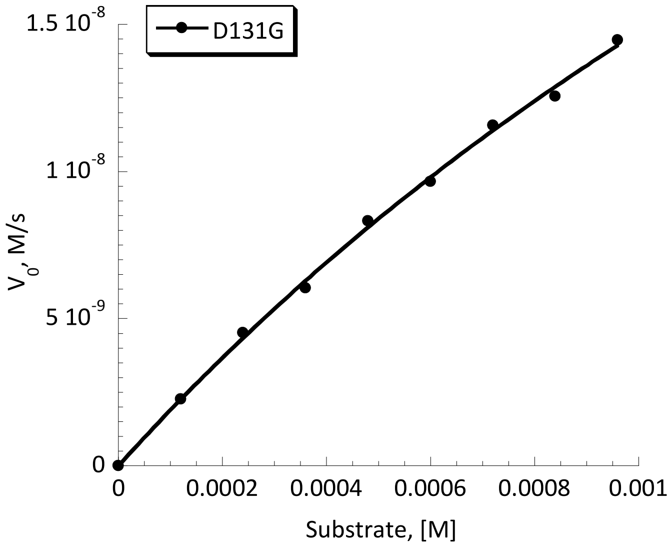


Figure S.01 Kemp Elimination activity of D131G CaM F92E 7th R mutant fit to Michaelis-Menten equation pH 7.0, 0.09 μM protein. (Unpublished, TD10621).

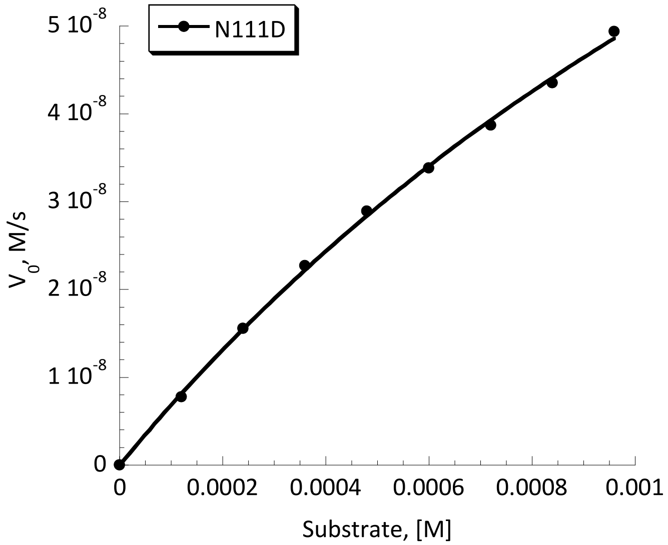


Figure S.02 Kemp Elimination activity of N111D CaM F92E 7th R mutant fit to Michaelis-Menten equation pH 7.0, 0.10 μM protein. (Unpublished, TD1050)

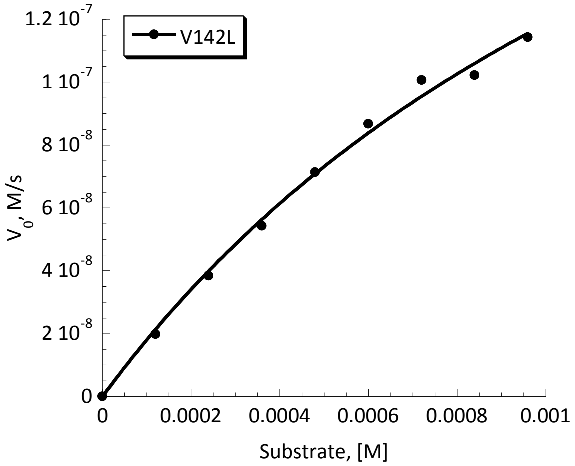


Figure S.03 Kemp Elimination activity of V142L CaM F92E 7th R mutant fit to Michaelis-Menten equation pH 7.0, 0.21 μM protein. (Unpublished, TD10543).

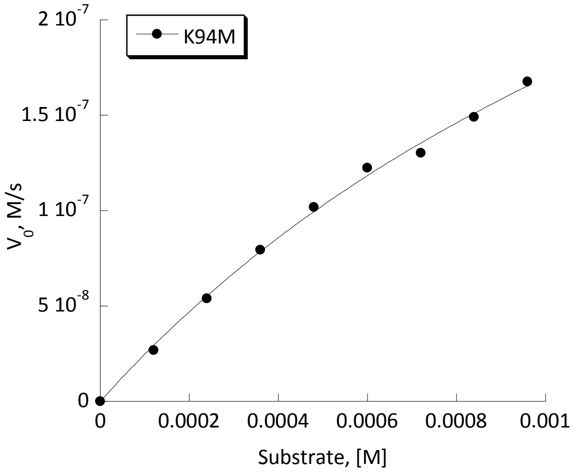


Figure S.04 Kemp Elimination activity of K94M CaM F92E 7th R mutant fit to Michaelis-Menten equation pH 8.0, 0.20 μM protein. (Unpublished, TD10622).

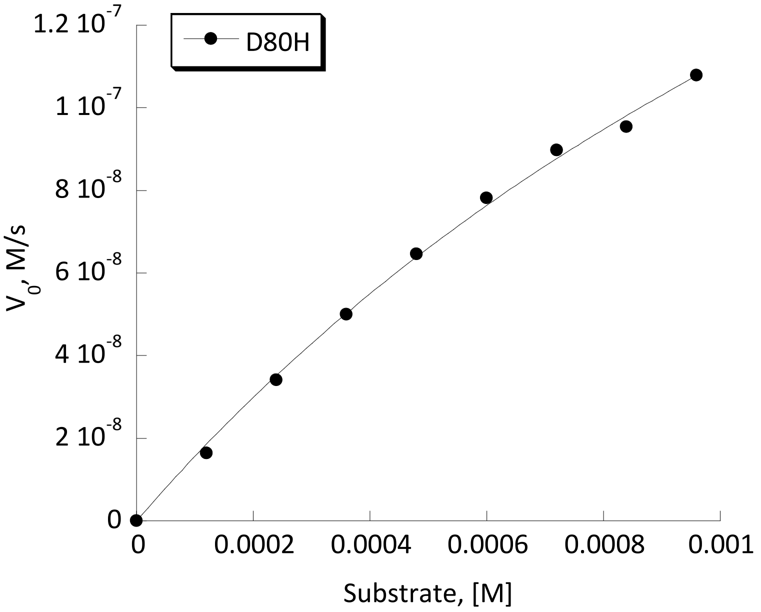


Figure S.05 Kemp Elimination activity of D80H CaM F92E 7th R mutant fit to Michaelis-Menten equation pH 7.0, 0.20 μM protein. (Unpublished, TD10652).

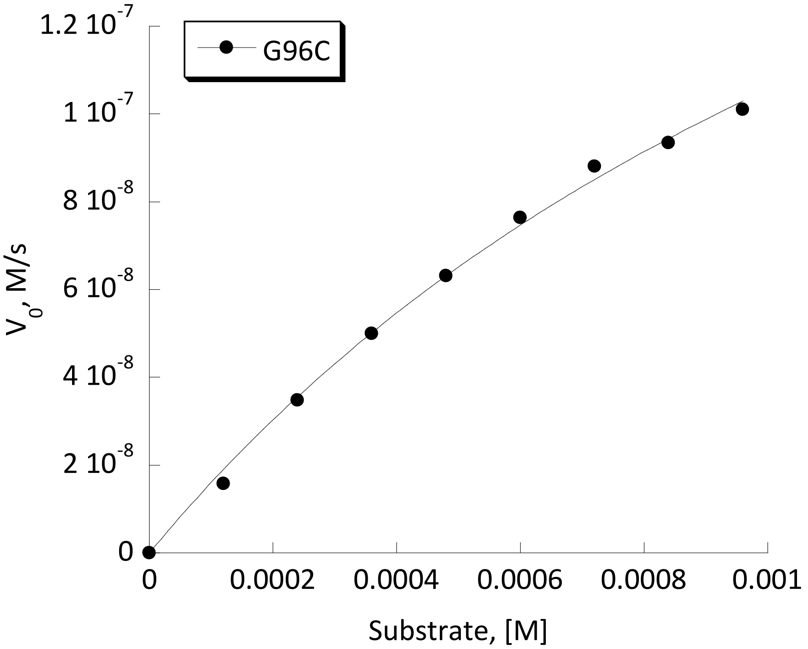


Figure S.06 Kemp Elimination activity of G96C CaM F92E 7th R mutant fit to Michaelis-Menten equation pH 7.0, 0.28 μM protein. (Unpublished, TD10652).

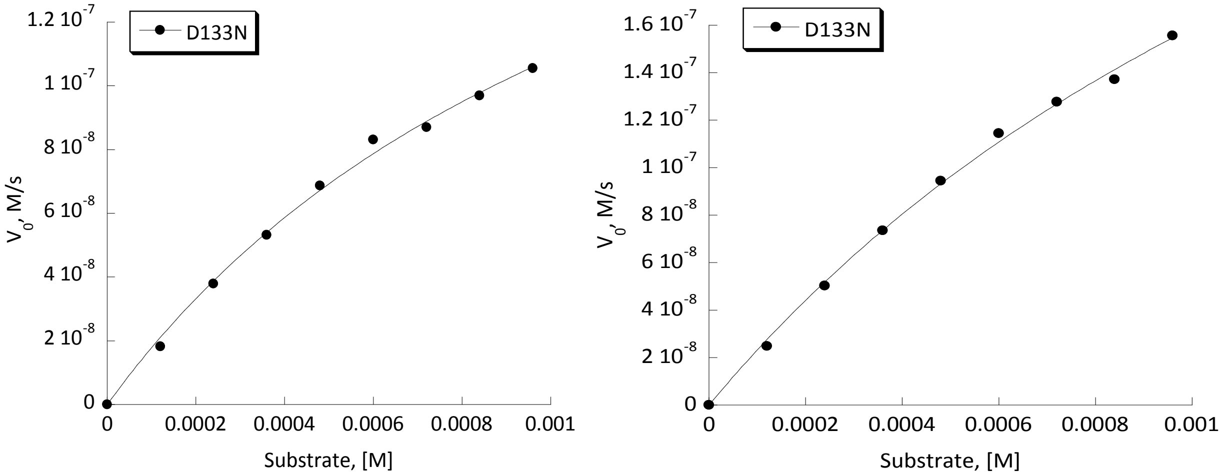


Figure S.07 Kemp Elimination activity of D133N CaM F92E 7th R mutant fit to Michaelis-Menten equation pH 7.0 (left), pH 8.0 (right) 0.20 μM protein. (Unpublished, TD10652 and TD10622).

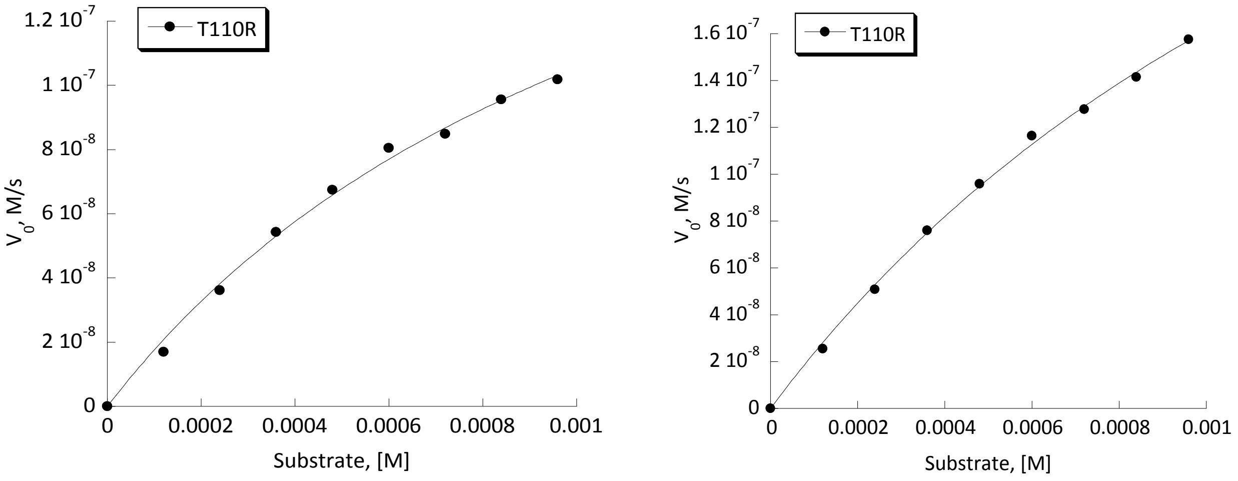
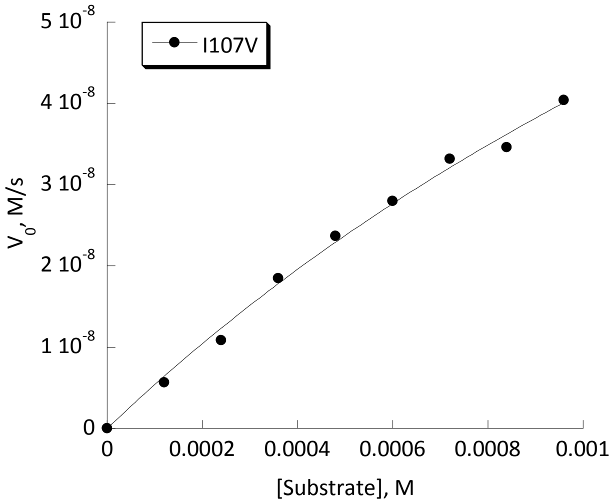


Figure S.08 Kemp Elimination activity of T110R CaM F92E 7th R mutant fit to Michaelis-Menten equation pH 7.0 (left), pH 8.0 (right) 0.21 μM protein. (Unpublished, TD10621 and TD10622).



S.09 Kemp Elimination activity of I107V CaM F92E 7th R mutant fit to Michaelis-Menten equation. (Unpublished, TD10353).

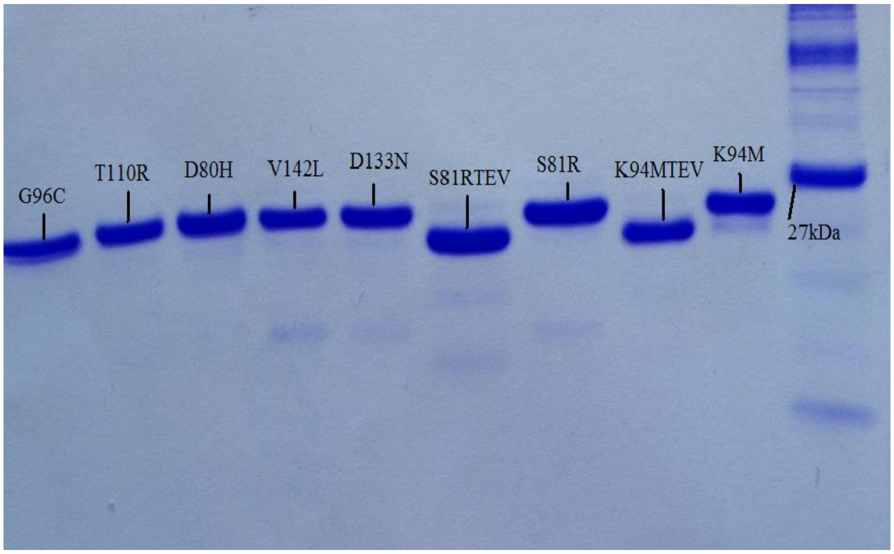

G96C
T110R
D80H
V142L
D133N
S81RTEV
S81R
K94MTEV
K94M
27kDa

Figure S.010. 10% SDS-PAGE CaM F92E 7thR mutants stained with coomassie blue. S81R and K94M have tags removed by TEV protease. (Unpublished, TD10753).

3. Site directed Mutagenesis of a non-enzymatic protein for a natural reaction

All work in the following chapter has been performed by the following authors unless otherwise stipulated

Work from this chapter has led in part to the following publication.

Moroz, Y. S., **Dunston, T. T**., Makhlynets, O. V., Moroz, O. V., Wu, Y., Yoon, J. H., Olsen, A. B., McLaughlin, J. M., Mack, K. L., Gosavi, P. M., van Nuland, N. A. J., and Korendovych, I. V. "New Tricks for Old Proteins: Single Mutations in a Nonenzymatic Protein Give Rise to Various Enzymatic Activities." *Journal of the American Chemical Society*. **2015** 137(47): 14905-14911.

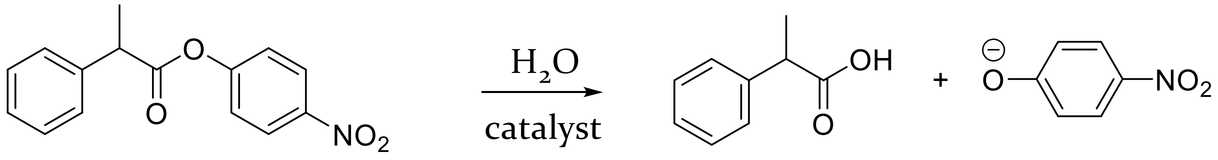
3.1 Introduction

In the previous chapter, we've showed that a small C-terminal fragment of calmodulin can catalyze an unnatural reaction with a 200 fold increase in catalytic activity. Since calmodulin is such a versatile and promiscuous protein with great flexibility, we were interested in expanding the reaction scope for catalyzing other reactions, specifically a natural reaction. The question becomes: Can we use calmodulin to design a catalyst for ester hydrolysis using a minimalistic approach? Specifically we are interested in hydrolysis of para-nitro phenyl ester substrates.[79,80-81]

There are three main methods used for protein structure prediction and for generating functional enzymes which are theozyme, iterative, and minimalist based approaches.[82-83] Theozyme based designs utilizes theoretical enzymes as an array of functional groups for transition state stabilization with the help of quantum mechanics and complex algorithms. Iterative based approaches predicts proteins functions iteratively using algorithms for calculating protein-protein function similarities. Using minimalistic approach focuses on the bare minimum requirements to achieve activity. By utilizing basic physico-chemical properties and strategic placing of only few highly active residues one can feasibly sample a very large variety of possible catalysts. In more general terms, minimalist approach looks for the mere *possibility of catalysis*, rather than trying to identify the *most active catalyst possible*. This approach provides the entry way for subsequent evolution.

Many proteins today have been designed using one or a combination of all protein design strategies. Baker *et al.* created an esterase for hydrolyzing para nitro phenol esters using theozyme based algorithms.[84] We decided to carry out the basic minimalistic route, where there are no complex calculations to set up and experimentally can be mastered in a very short time. Naturally occurring esterases rely on multiple amino acids arranged as dyads or triads stabilized by oxyanion hole for catalysis. It has been shown that a single histidine residue in a protein scaffold can catalyze ester hydrolysis with modest activity.[84] Ester hydrolysis is involved in many chemical processes such as the production of soap, wine, and the conversion of triglycerides to biodiesel.[85-86]

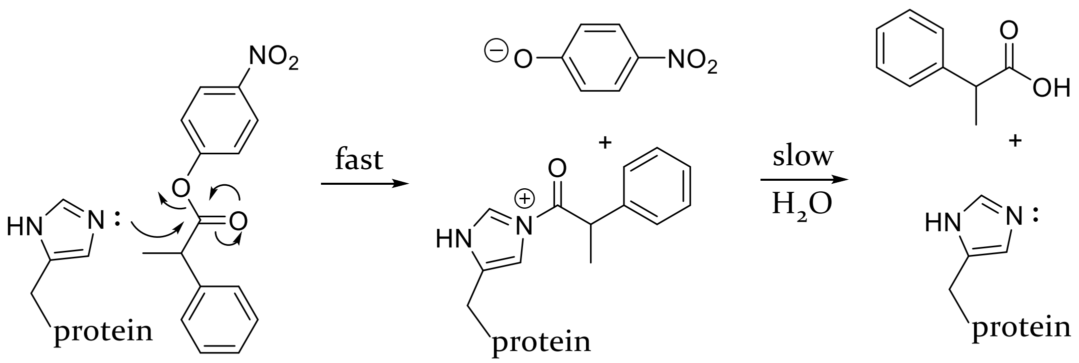
The target ester substrate reaction is depicted in Figure 3.1. The substrate p-nitrophenyl (2-phenyl)-propanoate (pNPP) hydrolyzes to produce 2-phenyl propionic acid and the yellow colored product para-nitrophenolate that can be easily observed and monitored spectroscopically at 405 nm in aqueous solution. pNPP is a well benchmarked ester substrate[23, 84] and advantageous for comparing to previous protein designs.



pNPP

Figure 3.1 Hydrolysis of p-nitrophenyl (2-phenyl)-propanoate (pNPP).

Our target reaction of interest is histidine mediated catalysis where histidine does a nucleophilic attack on the carbonyl carbon of the ester substrate to generate a high energy intermediate, which then collapse to form the product para-nitro phenol and regeneration of the catalyst (Figure 3.2). Substrates para-nitrophenyl acetate and para-nitrophenyl priopioniate are suitable chromogenic substrates commonly employed to characterize catalysts for ester hydrolysis.[79] A typical kinetic scheme of the hydrolysis of these substrates using a histidine nucleophile is outlined below.



$$E + S \overset{K_s}{\rightleftharpoons} ES \xrightarrow{k_2} ES' + P_1 \xrightarrow{k_3} E + P_2$$

Figure 3.2 Histidine mediated hydrolysis of p-nitrophenyl esters. Reaction demonstrating "classic" burst phase kinetics.

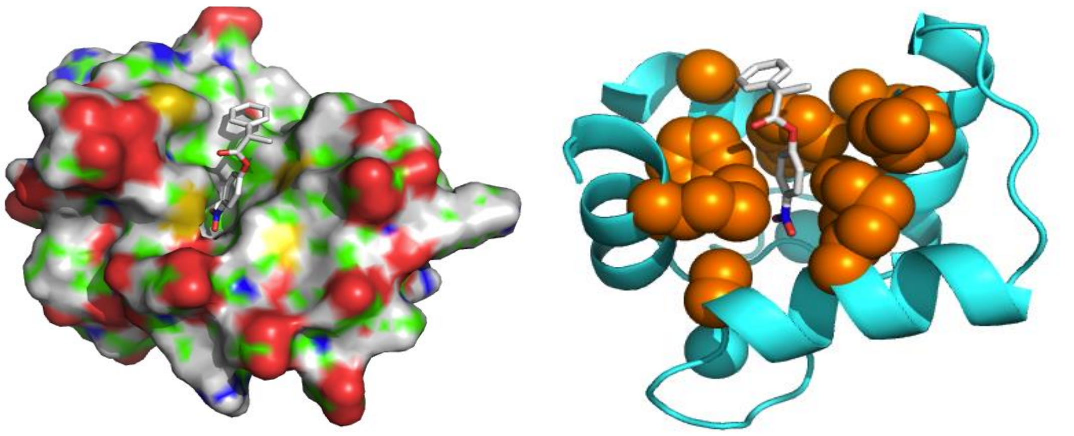
In the general kinetic scheme above, K_s represents the pseudo binding constant, k_2 is the rate of acylation (rate determining step) of the activated transition state, and k_3 is the rate of deacylation.

3.2 Design and characterization of a robust efficient esterase

3.2.1 Site-directed mutagenesis: Initial computational strategy

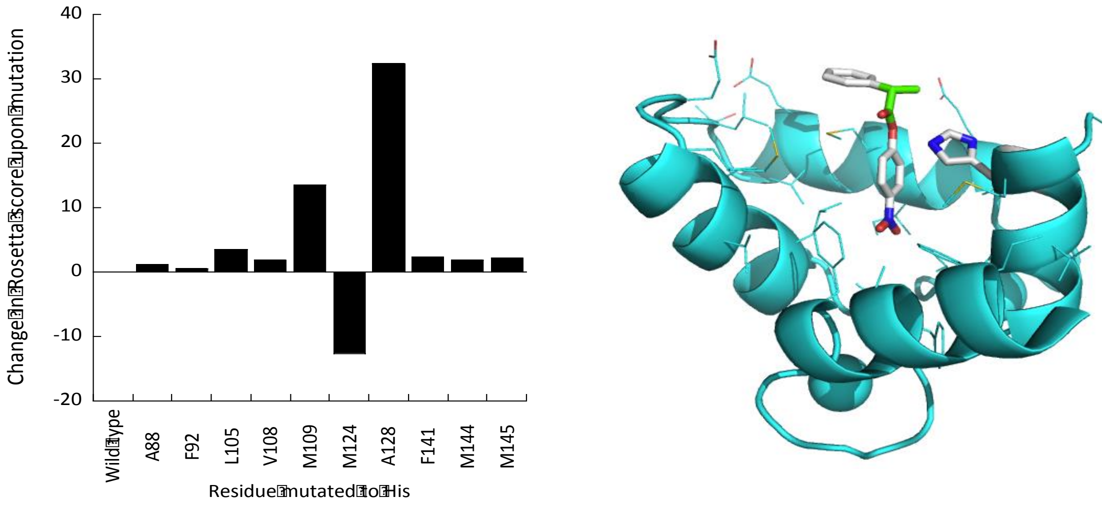
Directed evolution coupled with computational enzyme design has been known to enhance catalytic efficiencies in enzymes. Therefore, we used site directed mutagenesis to specifically and strategically introduce a histidine residue in calmodulin for catalyzing ester hydrolysis. An efficient catalyst should be small in size, stable, easily obtained, and allosterically regulated. Most natural enzymes are not allosterically regulated, so this requirement for our design would allow great switchable properties in the presence of calcium. Now that we know the requirements, the next question is how can we obtain the catalyst? As mentioned previously, we need an active site nucleophile, which is histidine and a substrate binding site. Lastly, allosteric control over the enzyme's function is desirable.

In a process of the minimalistic design strategy, we must undertake four steps (Figure 3.3). First, it is important to ensure that the substrate can bind to CaM. Secondly, using Autodock Vina software[87], the residues facing the docked substrate of calmodulin were identified that can serve as potential catalysts. Thirdly, each of those identified "hot spots" was computationally mutated to histidine and using Rosetta protein structure prediction program, we were able to thermodynamically score the protein's ability to tolerate those mutations. The higher the score (the difference between the Rosetta score of the native side chain and the score of the introduced amino acid), the less favorable that mutation is for being tolerated. The mutation with the lowest energy rotamer is the most favorable. Finally, active site transition state geometry must be tested to ensure proximity between enzyme and substrate when catalyzing the reaction. Although position 124 could accommodate histidine mutation, position 144 was the only that could satisfy all the above-mentioned requirements. This M144H variant was nicknamed AlleyCatE, by analogy to our previous protein AlleyCat designs, representative of only C-terminal domain of calmodulin and "E" for esterase.



Step 1
Ensure that the substrate
may bind to the protein

Step 2
Identify potential sites (orange)
for the active residue



Step 3
Check whether the mutation can
be accommodated by the protein

Step 4
Test whether transition state
geometry is possible

Figure 3.3 Computational methodology employed for identifying desired variant for ester hydrolysis reaction. Potential sites for introduction of histidine are outlined in orange.

3.2.2 Crude Cell Lysate Screening

To ensure that AlleyCatE is evolvable, directed evolution studies were performed by monitoring the ester hydrolysis reaction in crude cell lysates. Bacteria can naturally produce enzymes such as proteases and lipases that can catalyze the reaction as well, therefore *E. coli* pLysS was used as a control for the background reaction due to high efficiency of protein expression.

High throughput screening can be done in crude cell lysates allowing us to screen thousands of desired variants with efficient catalytic activity in a short time period.[38, 88-89] Therefore, applying this technique, we sought to examine if AlleyCatE is active above the background rate. The initial rates were monitored by observing the change in absorbance over time at 405 nm for AlleyCatE, full length M144H, the original parent calmodulin without M144H mutation, and *E. coli* (Figure 3.4). For those scaffolds catalyzing the reaction, a slight yellow color appears for the formation of product.

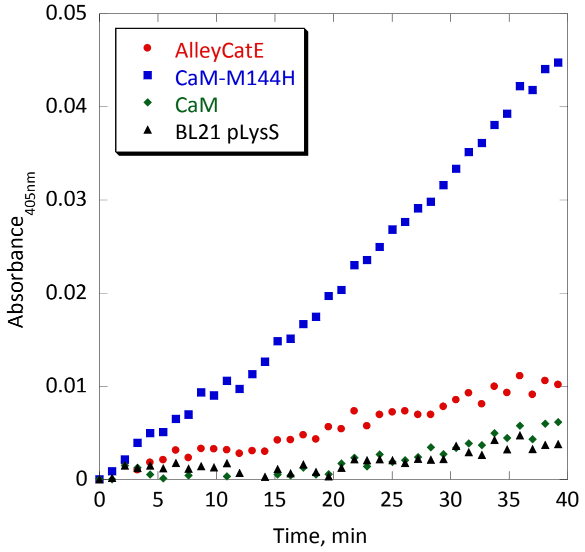


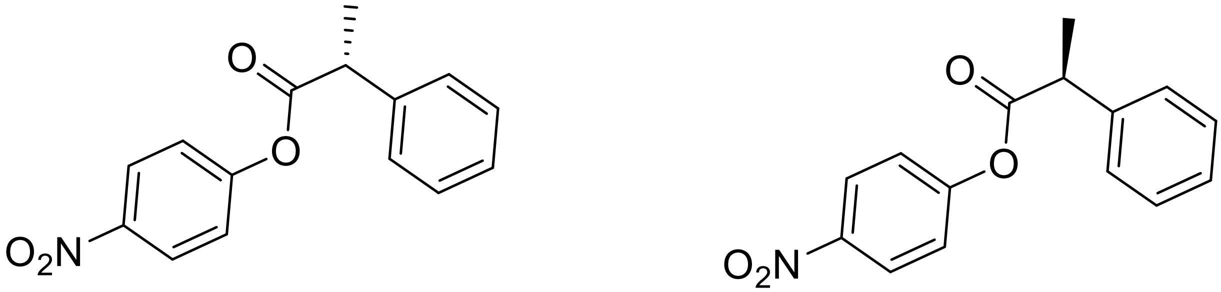
Figure 3.4 pNPP hydrolysis by crude cell lysates of *E. coli*: BL21 (DE3) pLysS cells (black triangles); *E. coli* BL21(DE3) pLysS cells containing a plasmid encoding c-CaM gene (green diamonds); *E. coli* BL21(DE3) pLysS cells containing a plasmid encoding AlleyCatE (red circles); and *E. coli* BL21(DE3) pLysS cells containing a plasmid encoding full length CaM M144H (blue squares). Conditions: 20 mM HEPES (pH 8.0); 100 mM NaCl; 10 mM CaCl2; 0.2% Triton X-100; 25 µM pNPP. Absorbance was monitored at 405 nm.

Although, the full length calmodulin shows more than 2 fold increase in activity above the background rate, AlleyCatE also shows a drastically higher increase. In figure 3.4, when the proteins are over expressed in *E. coli*, the rate of the reaction is observed higher than the background reaction. This is shown for both CaM M144H and AlleyCatE. Therefore, we were able to subject AlleyCatE to directed evolution. Additional A128T mutation provided extra increase in the esterase activity. This new double variant is nicknamed AlleyCatE2.

Published work was done with AlleyCatE, however much research was done employing full length calmodulin as a starting scaffold for evolution (discussed later in section 3.2.8).

3.2.3 Exploring substrate recognition and stereospecficity of 4-nitro-2-phenyl propanoate and para-nitro phenyl acetate

Ligand docking showed that the R isomer of pNPP binds tighter in the active site of AlleyCatE.[87] Since natural enzymes are often stereospecific, AlleyCatE and AlleyCatE2 enantioselectivity towards R, S and racemic pNPP was first explored and characterized by Michaelis-Menten kinetics. Structures of R and S ester substrates were synthesized by Dr. Yurii Moroz and are displayed in Figure 3.5.



R-pNPP S-pNPP

Figure 3.5 R and S isomers of pNPP

We have characterized kinetic traces of R-pNPP, S-pNPP, and pNPA hydrolysis using the method developed by Kezdy and Bender[90] and fit to the equation below:

$$[p - nitrophenolate] = At + B(1 - e^{-bt})$$

where *t* is time. Plotting reciprocal values of A and b versus $1/[S]_0$ at various initial substrate concentrations (see appendix for graphs) allows for straightforward determination of k_{cat}, K_M(app),k_2, and K_s.[90]

AlleyCatE and AlleyCatE2 were characterized fully. The k_2/K_s is nearly 2 fold higher and the k_{cat}/K_M is nearly 40% higher than those of AlleyCatE. Moreover, calmodulin's promiscuity to encapsulate different substrates was examined by probing the degree of substrate selectivity using pNPA. Hydrolysis of pNPA compared to pNPP were significantly lower (26-35 fold) (Table 3.1). These results demonstrate that using a minimalistic design approach, one can efficiently design an esterase with improved activity and provide a degree of substrate recognition.

Table 3.1 Kinetic parameters for AlleyCatE and AlleyCatE2.

Protein	Substrate	k_2/K_S, $M^{-1}\,min^{-1}$	$k_2 \times 10^{-2}$, min^{-1}	K_S, μM	k_{cat}/K_M, $M^{-1}\,min^{-1}$	$k_{cat} \times 10^{-2}$, min^{-1}	K_M (app) , μM
AlleyCatE	R-pNPP	70,000±17,000	140±45	22±6	4,800 ±300	8.8±0.7	18±1
	S-pNPP	15,000±200	140±40	86±26	1,700 ±200	3.7±0.4	21±3
	pNPA	1,950±170	75±13	400±70	330 ±45	24±10	700 ±300
AlleyCatE2	R-pNPP	120,000±27,000	160±30	13±3	6,600 ± 600	28 ±1	42 ±4
	S-pNPP	26,400±1,000	220±30	83±11	2,400±100	6.5±0.3	27 ± 2
	pNPA	4,600±500	90±9	200±24	360 ± 30	7.0±0.5	200 ±15

The raw data used for the fits are presented in Figs 3.6-3.8.

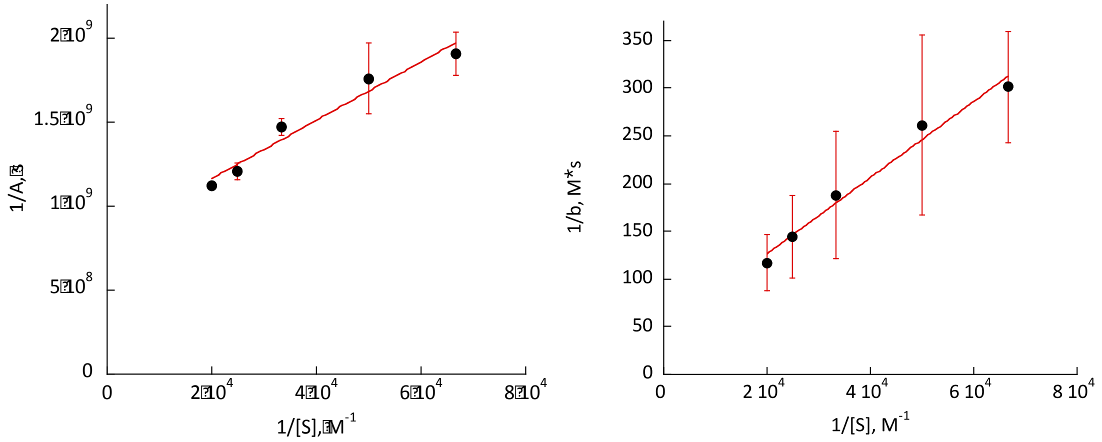


Figure 3.6a Analysis of the kinetic data for hydrolysis of S-pNPP catalyzed by AlleyCatE.

Conditions: 20 mM HEPES (pH 7.5); 100 mM NaCl; 10 mM $CaCl_2$; 2.0 μM AlleyCatE.

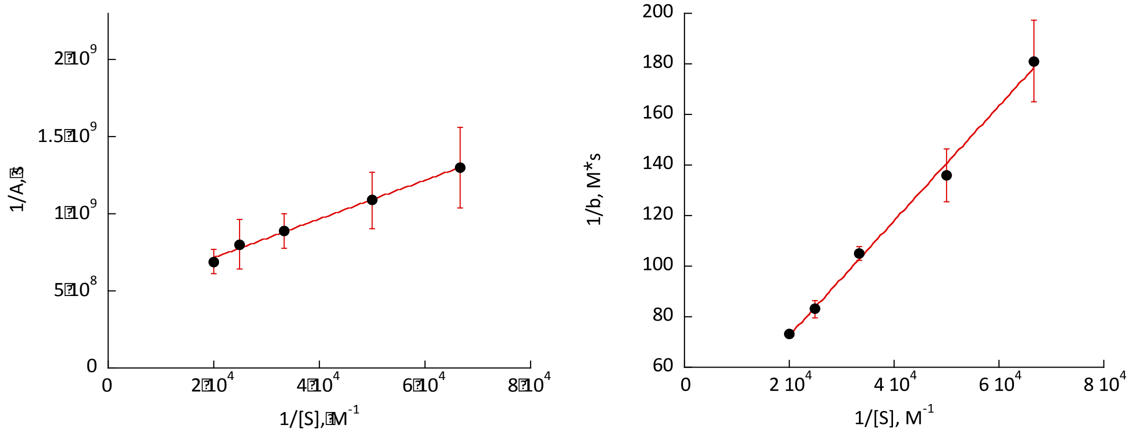


Figure 3.6b Analysis of the kinetic data for hydrolysis of S-pNPP catalyzed by AlleyCatE2.

Conditions: 20 mM HEPES (pH 7.5); 100 mM NaCl; 10 mM $CaCl_2$; 2.0 μM AlleyCatE2.

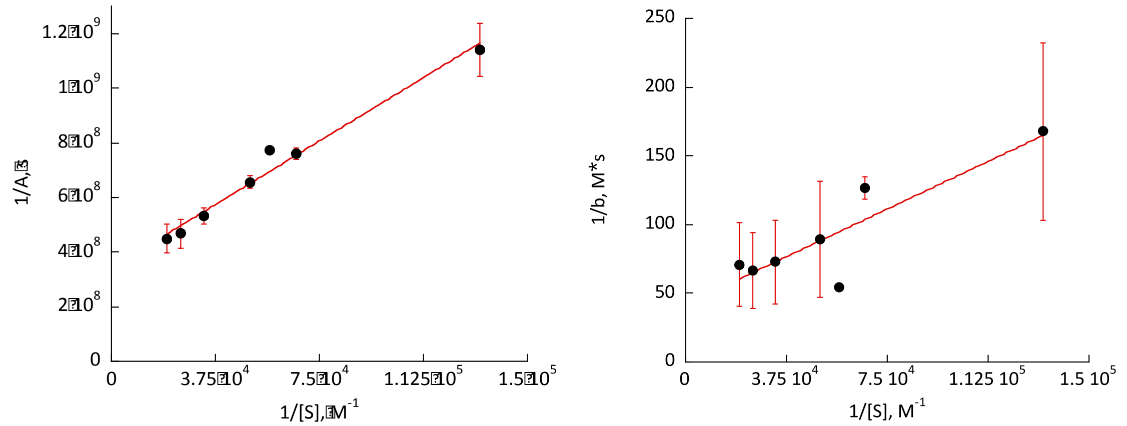


Figure 3.7a Analysis of the kinetic data for hydrolysis of R-pNPP catalyzed by AlleyCatE.

Conditions: 20 mM HEPES (pH 7.5); 100 mM NaCl; 10 mM $CaCl_2$; 2.0 µM AlleyCatE.

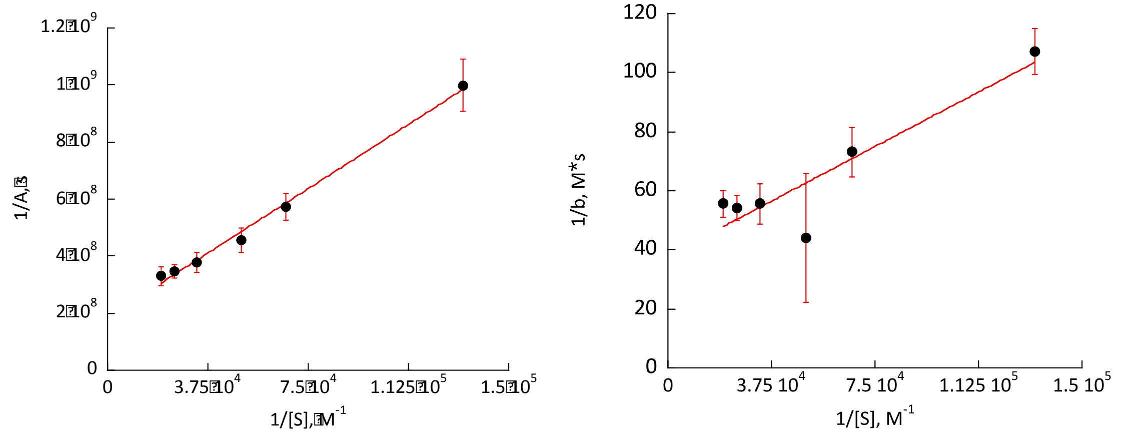


Figure 3.7b Analysis of the kinetic data for hydrolysis of R-pNPP catalyzed by AlleyCatE2.

Conditions: 20 mM HEPES (pH 7.5); 100 mM NaCl; 10 mM $CaCl_2$; 2.0 µM AlleyCatE2.

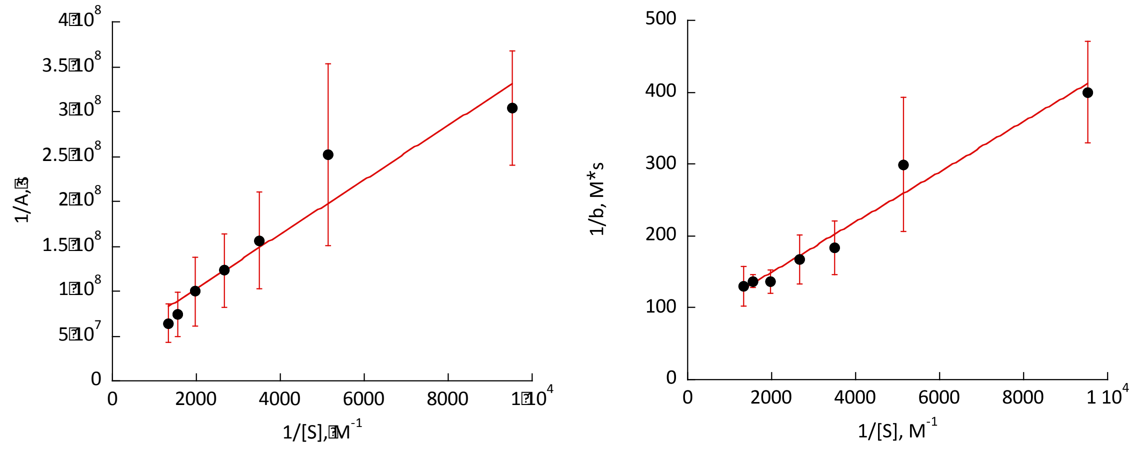


Figure 3.8a Analysis of the kinetic data for hydrolysis of pNPA catalyzed by AlleyCatE.

Conditions: 20 mM HEPES (pH 7.5); 100 mM NaCl; 10 mM $CaCl_2$; 6.0 μM AlleyCatE.

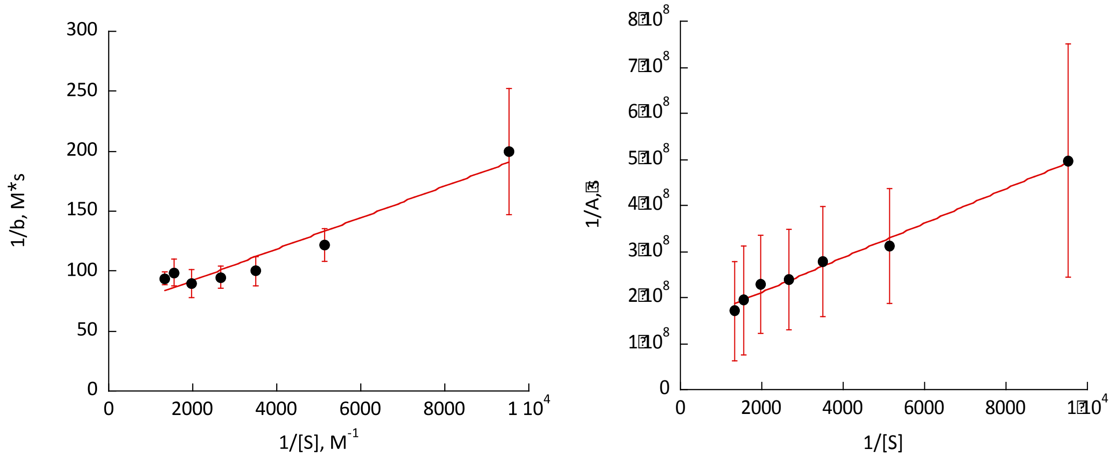


Figure 3.8b Analysis of the kinetic data for hydrolysis of pNPA catalyzed by AlleyCatE2.

Conditions: 20 mM HEPES (pH 7.5); 100 mM NaCl; 10 mM $CaCl_2$; 6.0 μM AlleyCatE2.

3.2.4 Protein Folding: Conformational Stability

Proteins folds are stabilized by their secondary structures. Introducing perturbations into a protein's structure can either increase or decrease the catalytic activity. Circular dichroism was done first on full length and c-terminal constructs of M144H (AlleyCatE) to observe the degree of α-helicity and from that we saw that AlleyCatE was able to maintain native calmodulin's α-helical shape although with less α-helical content.

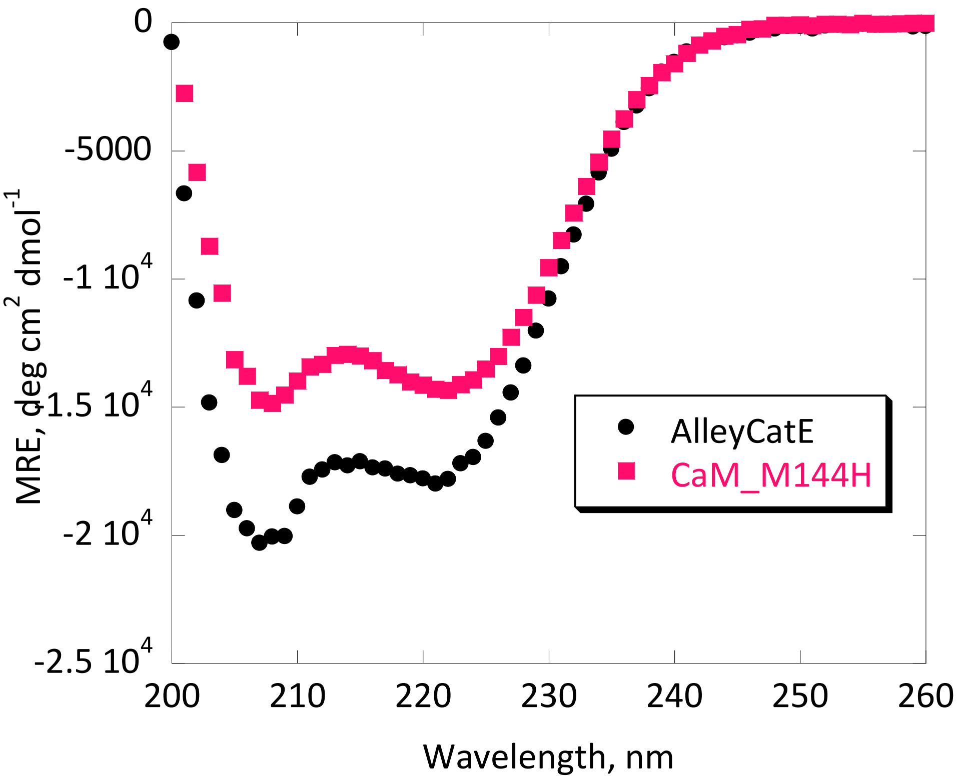


Figure 3.9 CD spectrum of AlleyCatE and CaM-M144H. Conditions: 25.0 μM protein, 4 mM HEPES (pH 7.5), 2 mM CaCl2 and 30 mM NaCl.

To investigate whether introducing histidine into calmodulin can perturb the protein's overall fold, a denaturation experiment was done in the presence of 0-6 M guanidine hydrochloride. From this experiment, the free energy of protein folding and unfolding was assessed. As seen in Figure 3.10, at very low concentrations of GdmCl both wild type C-terminal calmodulin and AlleyCatE remain folded until about 4M where both transition into the unfolded protein states (Figure 3.10). The M144H mutation in AlleyCatE only results in a small loss of the free energy of folding which was well-tolerated by the highly stable calmodulin fold.[27] Therefore, the introduction of histidine into C-terminal calmodulin did not substantially disrupt the protein's overall fold. These results demonstrates the conformational stability maintained with a single histidine mutation. Calculated free energies for both proteins are displayed in Table 3.2.

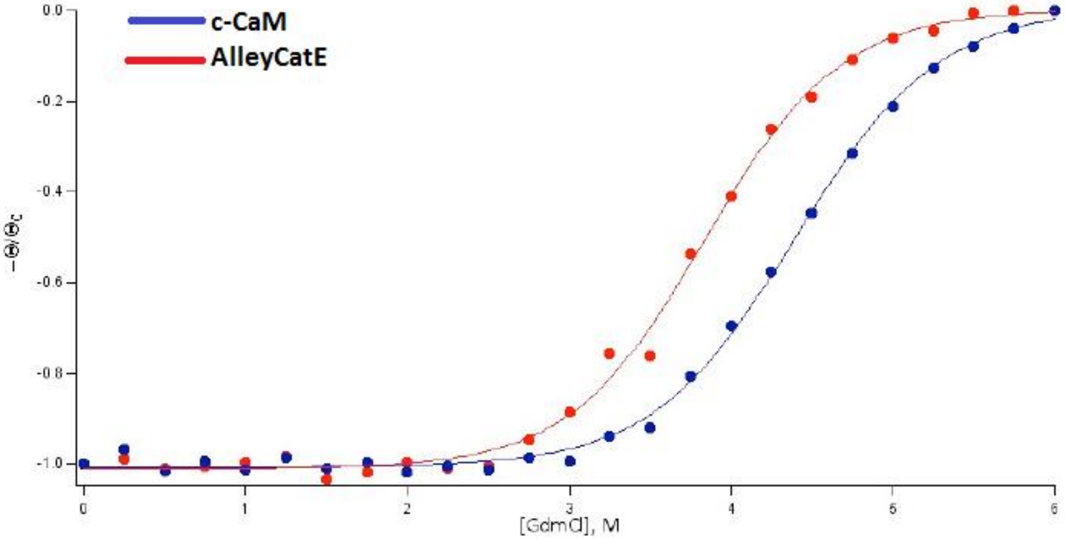


Figure 3.10 Chemical denaturation (guanidium hydrochloride GdmCl) profiles of C-CaM (blue circles) and AlleyCatE (red circles).

Table 3.2 Thermodynamic parameters obtained from denaturation studies.

Proteins	ΔG kJ/mol	m kJ/mol*M
c-CaM	24 ± 0.9	5.5 ± 0.2
AlleyCatE	22 ± 1.2	5.9 ± 0.3

3.2.5 MALDI-TOF for Capturing Acylated Intermediate

The hydrolysis of pNPP and pNPA proceeds by an initial burst phase mechanism to generate an acylated intermediate where the acyl group from the ester substrates are attached to the nitrogen from histidine on the protein. Initial studies were done on capturing the acylated transition state by MALDI-TOF spectrometry to ensure that the reaction occurred. Free AlleyCatE and AlleyCatE2 have masses of 8613 Da and 8643 Da respectively. When acylated by R-pNPP, AlleyCatE and AlleyCatE2 show masses of 8747 Da and 8777 Da respectively. As a control, C-CaM was done with R-pNPP as well producing no observable change in mass. It is clear that both AlleyCatE and AlleyCatE2 gets acylated during the reaction. This confirms that the reaction takes place and that the [ES] complex has formed (Figure 3.11). For C-CaM with R-pNPP the presence of an acylated peak is absent which shows that the acylation is not a random event and it further confirms that there is no reaction with C-CaM (Figure 3.12).

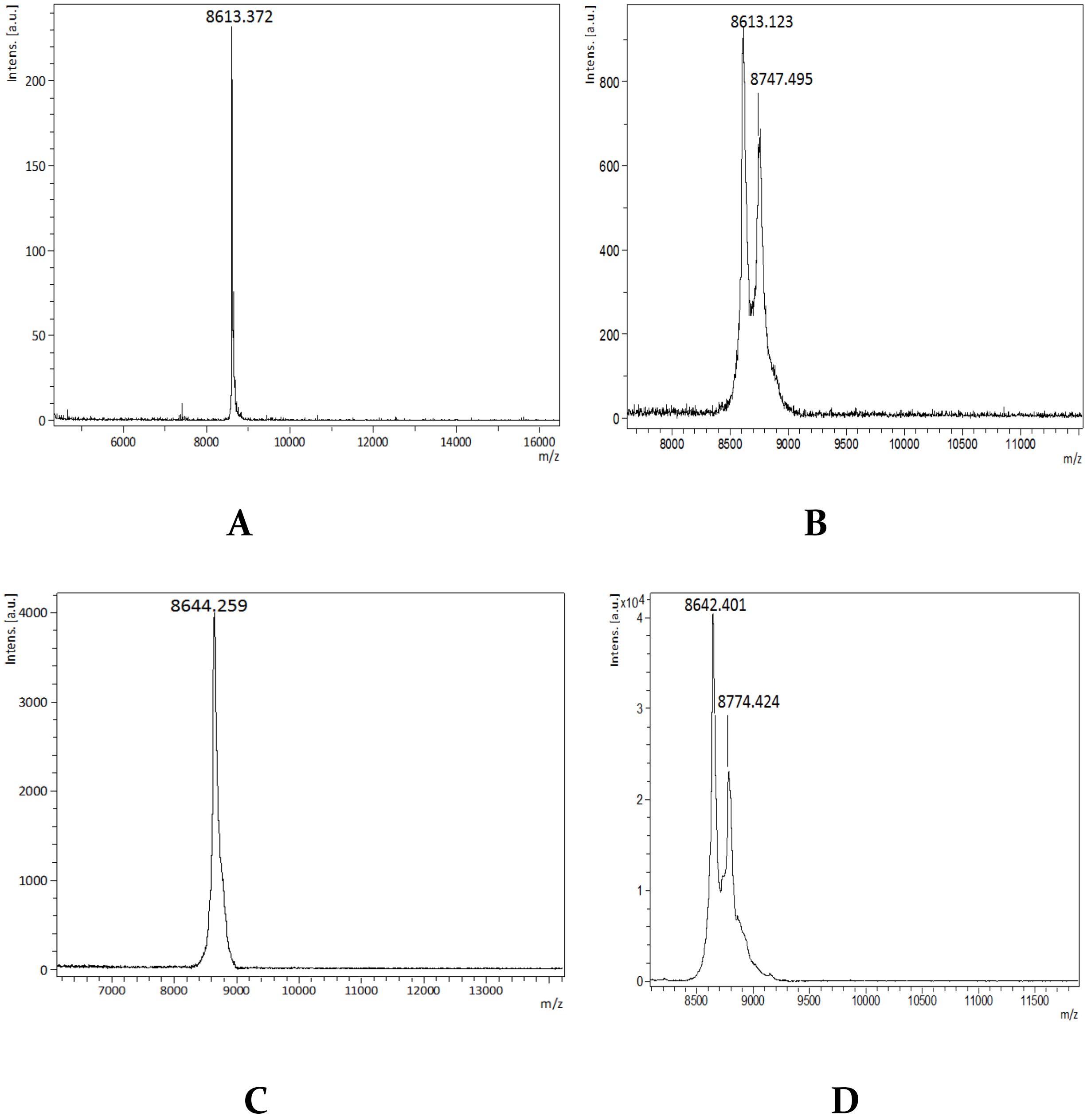


Figure 3.11 MALDI-TOF spectra of A: free AlleyCat E, B: AlleyCatE with 0.5 equiv.pNPP, C: free AleyCatE2, D: AlleyCatE2 with 0.5 equiv pNPP.

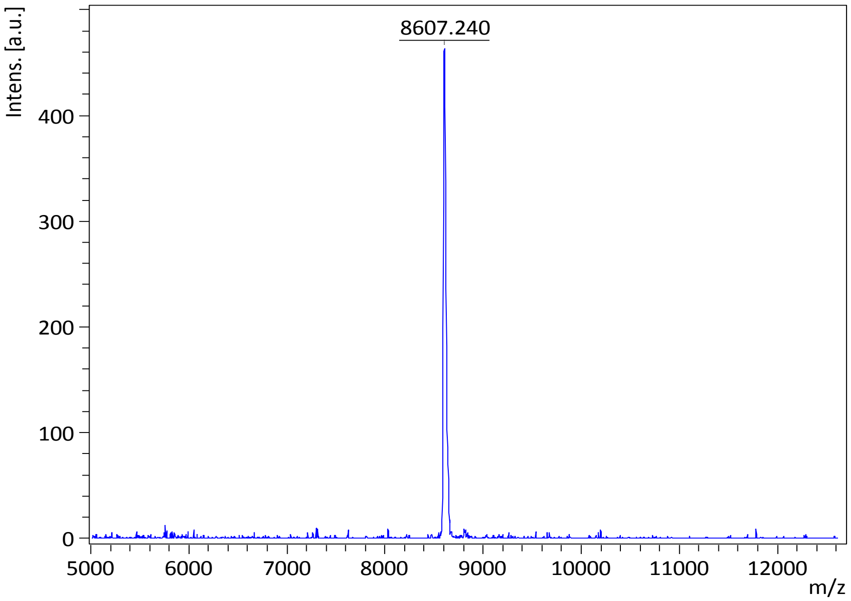


Figure 3.12 MALDI-TOF spectra of c-CaM in the presence of 2.5 equiv R-pNPP.

3.2.6 Assessing k_{cat} and product inhibition

For an enzyme to catalyze a reaction efficiently and successful it must be able to undergo several turnovers. The turnover number k_{cat} is defined as the maximum number of substrate molecules that an enzyme can convert into product per second.[77] In subsequent experiments consisted we monitor the turnover number (k_{cat}) of the reaction over time by UV-Vis spectrometry then by MALDI-TOF to test for completion of ester hydrolysis reaction. Measurements were taken over different time periods only for AlleyCatE to confirm reaction completion. As shown in Figure 3.13 AlleyCatE undergoes several acylation events over time. The k_{cat} was investigated for AlleyCatE over time. AlleyCatE undergoes at least 12 catalytic turnovers over 42 hours (Figure 3.14).

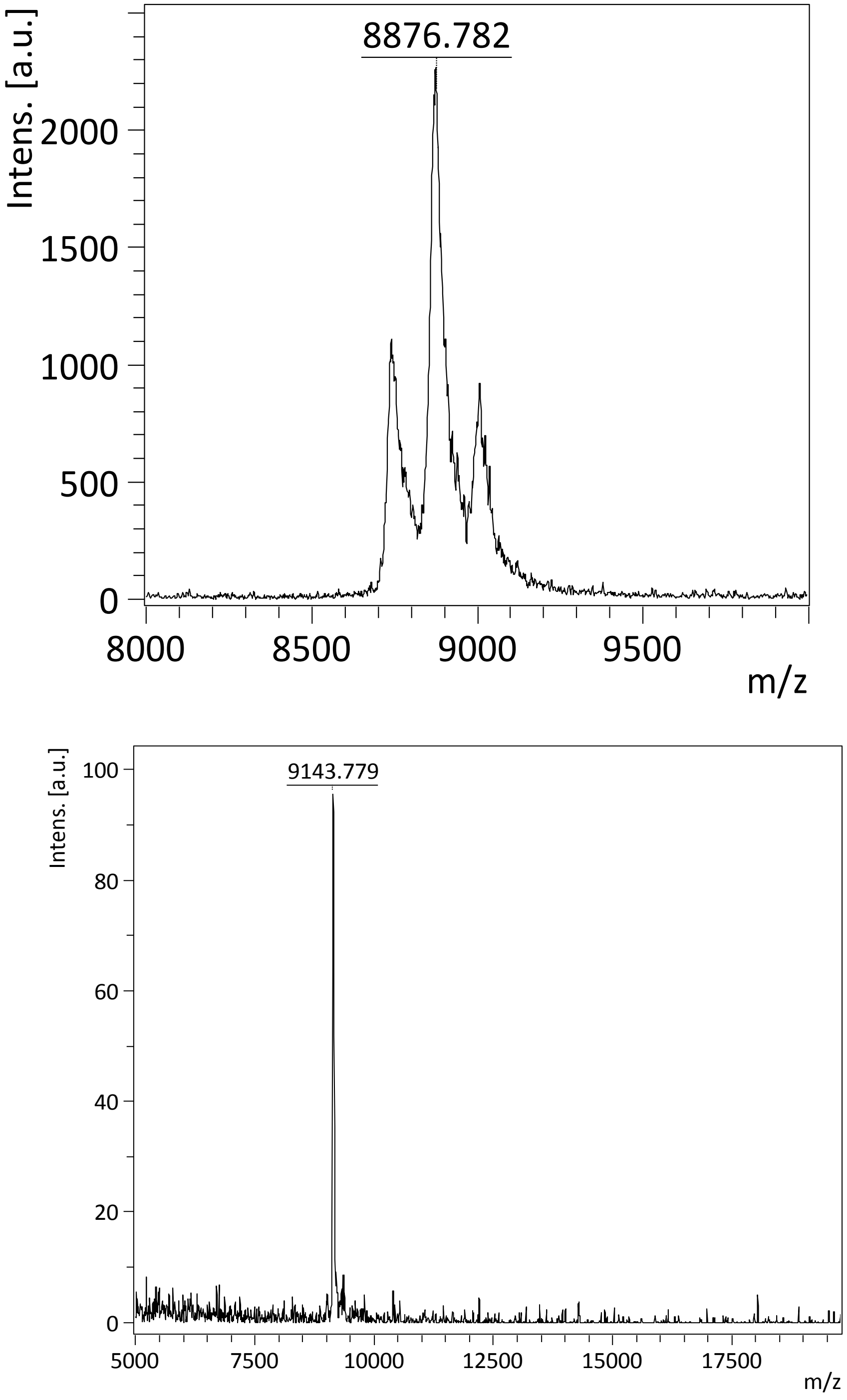


Figure 3.13 MALDI-TOF spectra of AlleyCatE at taken at different time points . Top: AlletCatE in the presence of 8 equivalents pNPP. The three peaks (zoomed in) shows

AlleyCatE single, double, and triply aclyated. Bottom: Full conversion of pNPP after 48 hours. The one peak shows AlleyCatE acylated four times.

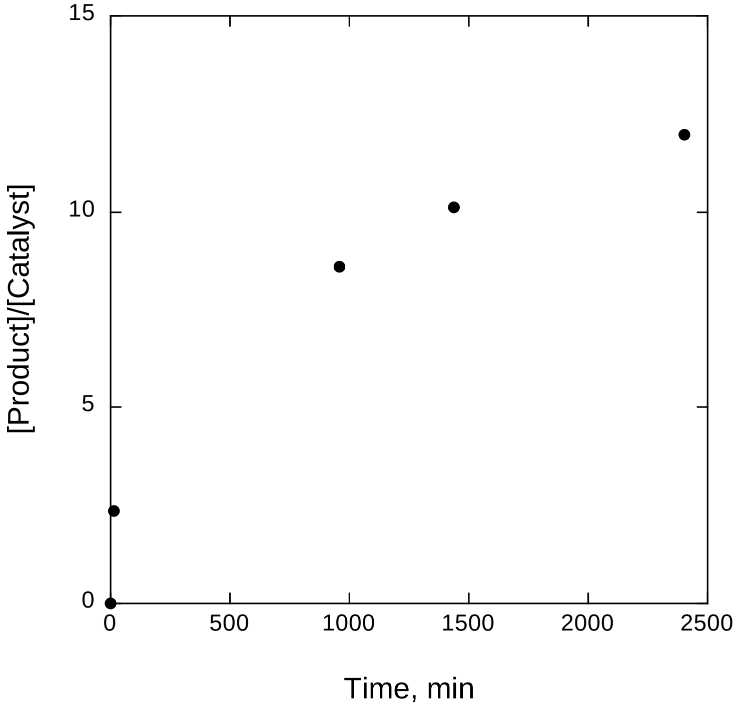


Figure 3.14 Substrate turnover by AlleyCatE. Conditions: 2.0 µM AlleyCatE, 60 µM R-pNPP in 20 mM HEPES, 100 mM NaCl, 10 mM $CaCl_2$, pH 7.5.

In enzyme kinetics, the products behave as inhibitors since they are naturally present or generated during the reaction mixture. Thus, they are termed product inhibitors and the phenomena is called product inhibition. Product inhibition studies were performed to ensure that ester hydrolysis of pNPP was not inhibited by 4-nitro phenol or 2-phenyl propionic acid (Figure 3.15). From the graph, it is clear that there is no product inhibition when adding either of the products to the reaction.

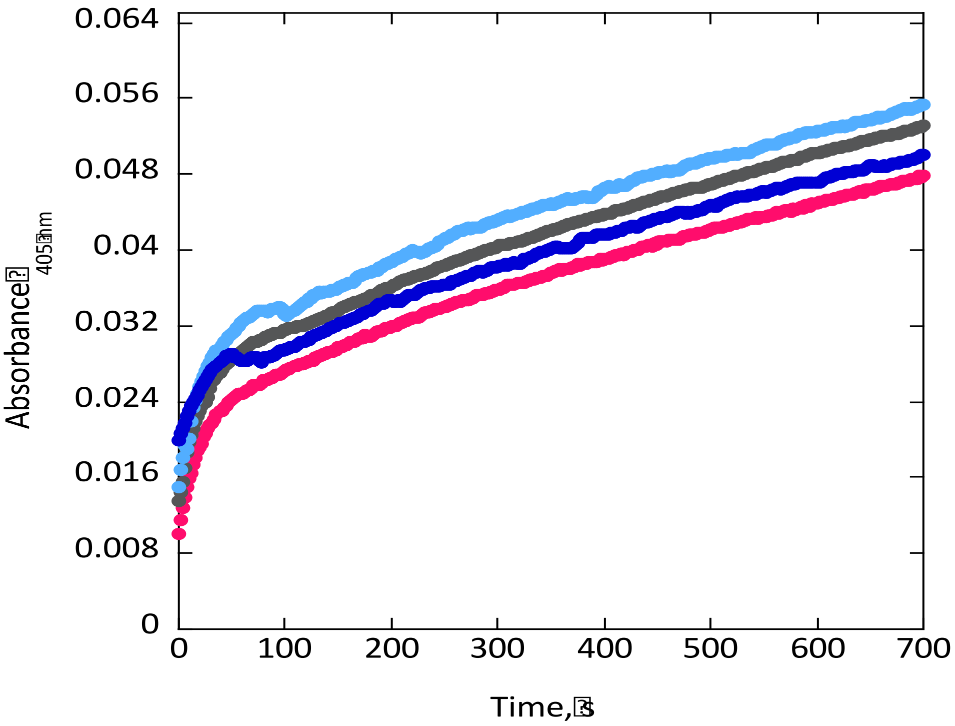


Figure 3.15 The photometric traces (at 405 nm) of R-pNPP hydrolysis catalyzed by AlleyCatE in the absence of product (black), in the presence of 1 equiv. of *p*-nitrophenol (light blue), in the presence of 5 equiv. of *p*-nitrophenol (dark blue), and in the presence of 5 equiv. of (R)-2-phenylpropionic acid (red). Conditions: 2.0 μM AlleyCatE, 60 μM R-pNPP in 20 mM HEPES, 100 mM NaCl, 10 mM $CaCl_2$, pH 7.5.

3.2.7 Calcium Allosteric Regulation Assay

Investigation of catalytic activity of ester hydrolysis in the presence and absence of calcium was assessed in a kinetic assay for AlleyCatE and AlleyCatE2. As previously mentioned, calmodulin adopts an open conformation upon binding calcium ions which gives rise to its promiscuous behavior. This behavior was characterized using pure protein samples with R-pNPP at 405 nm. AlleycatE and AlleyCatE2 are both at least 100-fold more active in the presence of calcium (Figure 3.16). Without calcium, the activity is significantly decreased and the proteins are no longer active, corresponding to calmodulin's apo state.

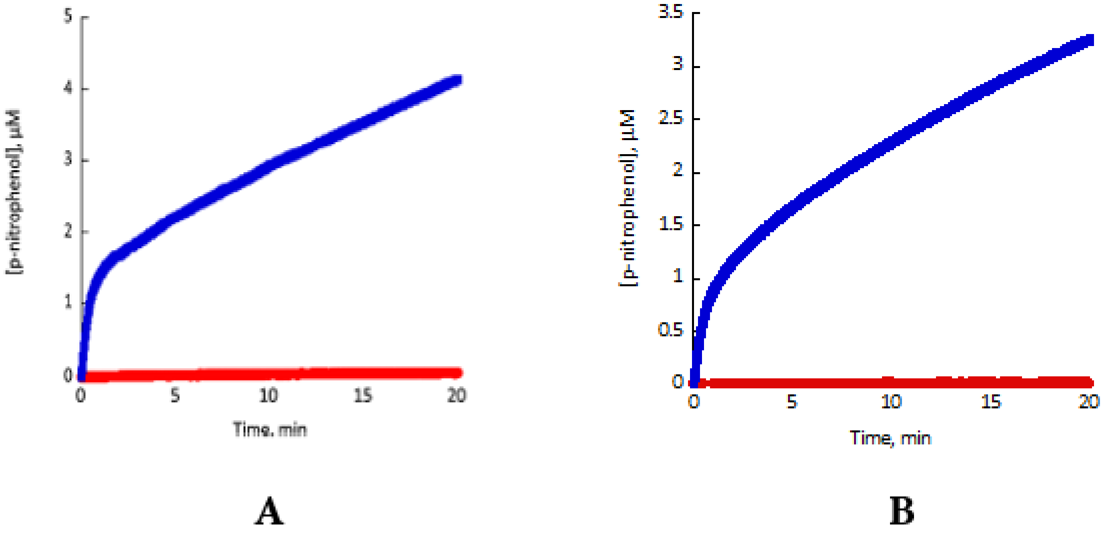


Figure 3.16 A). Hydrolysis of R-pNPP (50 µM) catalyzed by AlleyCatE (2 µM) in the presence (blue) and in the absence (red) of 10 mM CaCl2. B) Hydrolysis of R-pNPP (50 µM) catalyzed by AlleyCatE (2 µM) in the presence (blue) and in the absence (red) of 10 mM CaCl2. Conditions: 20 mM HEPES,100 mM NaCl, pH 7.5.

3.2.8 Characterization of other Full Length Esterases identified in the Saturation Mutagenesis work

Positions tested by saturation mutagenesis for full length M144H were D80, A88, F89, V91, L105, H107, V108, L112, and A128. The resulting variants obtained were V108G, A128T, A88G, A88T, and the wild type V108V (M144H).These proteins were tested for catalyzing ester hydrolysis of pNPP and pNPA substrates. Both burst phase and slow phase kinetics were examined on a platereader. The burst phase k_2/K_s for A128T, V108G, and wild- type V108V (M114H) were 160 $M^{-1}s^{-1}$, 111 $M^{-1}s^{-1}$, 111 $M^{-1}s^{-1}$ (Figure 3.17) and for slow phase activities were 37 $M^{-1}s^{-1}$, 23 $M^{-1}s^{-1}$, and 23 $M^{-1}s^{-1}$ respectively. For pNPA hydrolysis, A128T, V108G, and V108V (M144H) have burst phase catalytic efficiencies of 15 $M^{-1}s^{-1}$, 13 $M^{-1}s^{-1}$, and 11.2 $M^{-1}s^{-1}$ respectively and slow phase efficiencies of 3 $M^{-1}s^{-1}$, 8 $M^{-1}s^{-1}$, and 6 $M^{-1}s^{-1}$ (Figure 3.18). A128T mutant had the best activity for burst phase pNPP hydrolysis due to the much lower K_M. This result explains the introduction of A128T in the double variant AlleyCatE that was previously mentioned. The low k_{cat}/K_M for pNPA hydrolysis reveals low activity of these mutants for this substrate thus portraying substrate selectivity.

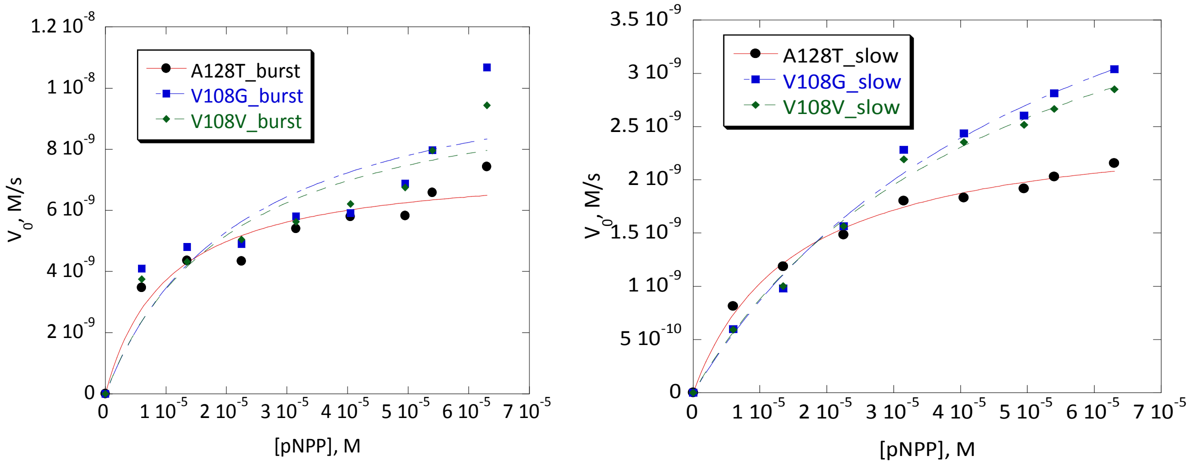


Figure 3.17 Left: pNPP burst phase kinetics of A128T, V108G, and wild type V108V (M144H) at pH 7.5. Right: pNPP slow phase reaction of A128T, V108G, and wild type V108V (M144H) at pH 7.5, 4.5 µM protein. (Unpublished, TD10951)

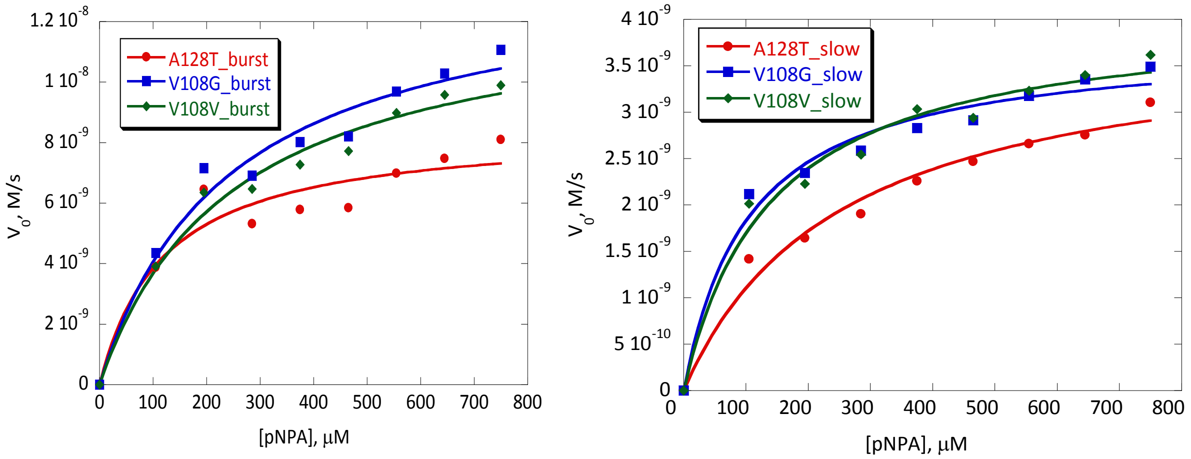


Figure 3.18 Left: pNPA burst phase kinetics of A128T, V108G, and wild type V108V (M144H) at pH 7.5. Right: pNPA slow phase reaction of A128T, V108G, and wild type V108V (M144H) at pH 7.5, 4.5 µM protein. (Unpublished, TD10941).

Slow phase pNPP hydrolysis was repeated for A128T, V108G, and wild-type V108V (M144H) for comparison with A88G and A88T (Figure 3.19). The proteins have the k_{cat}/K_M values of 36, 29, 25, 1, and 16 ($M^{-1}s^{-1}$) respectively. The only mutant that has a higher activity and lower K_M was A128T; therefore this mutant was further studied as a double variant with the second best mutant V108G. For slow phase k_{cat}/K_M was 26 $M^{-1}s^{-1}$ for V108GA128T and 24 $M^{-1}s^{-1}$ for wild-type M144H. For the burst phase, k_{cat}/K_M was 64 $M^{-1}s^{-1}$ for V108GA128T and 164 $M^{-1}s^{-1}$ for M144H. Considering only slow phase kinetics, V108GA128T has about the same catalytic activity as wild type, therefore it has not additionally improved the enzyme (Figure 3.20).

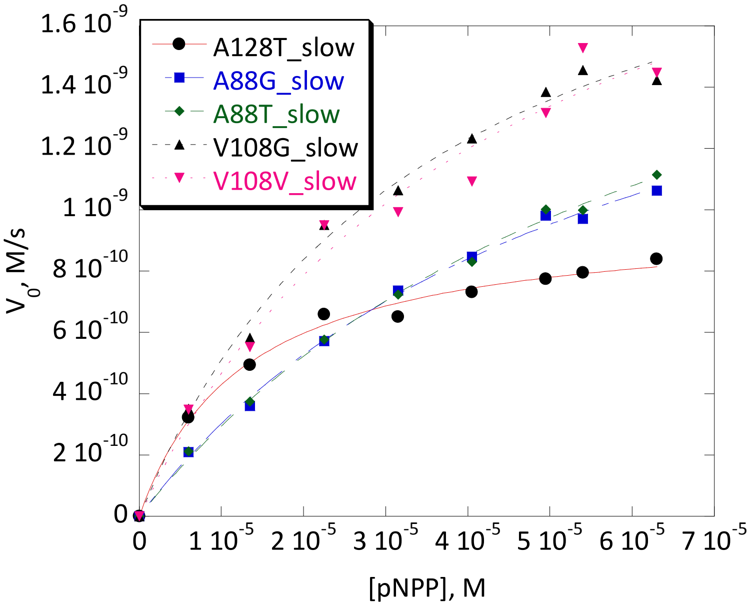


Figure 3.19 pNPP slow phase kinetics of A128T, V108G, and wild type V108V (M144H), A88G, and A88T at pH 7.5, 2.0 μM protein. (Unpublished, TD11021 and TD11091).

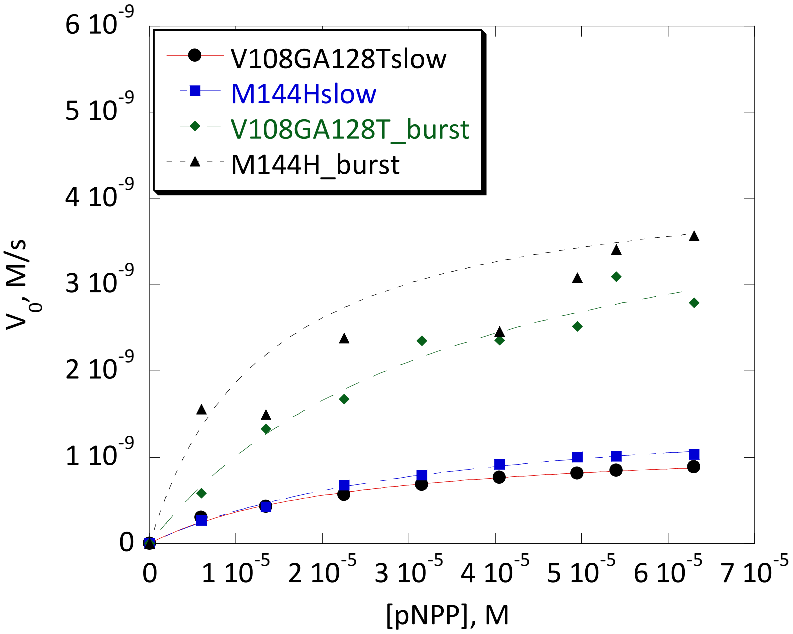


Figure 3.20 pNPP kinetics V108GA128T and wild type V108V (M144H at pH 7.5, 2.0 μM protein. (Unpublished, TD11091)

3.3 Conclusion

We have demonstrated how easily a minimalistic approach to protein design can be applied to different types of chemical reactions: introducing a *single* histidine residue into calmodulin confers esterase activity onto this non-enzymatic protein. Calmodulin's flexibility was demonstrated through fine tuning to increase stability, active site functionality, and catalytic efficiency. We were able to show that AlleycatE and AlleyCatE2 serve as efficient esterases for hydrolyzing pNPP and pNPA. All calmodulin esterase variants (full length and AlleyCat's) showed a degree of substrate recognition and selectivity for pNPP over pNPA esters. The obtained allosterically controlled esterase is active enough to compete against the background of natural enzymes in the *E. coli* crude lysate, which allowed for further activity improvement using directed evolution. We demonstrate that even a small 74-residue non-enzymatic protein can adapt different catalytic activities (Kemp elimination, ester hydrolysis) as a result of just a single mutation. These results are important for our understanding of how proteins evolve to adopt new functionality as well as how this knowledge could be applied to protein design.

Appendix Chapter 3 Supplemental Experiements

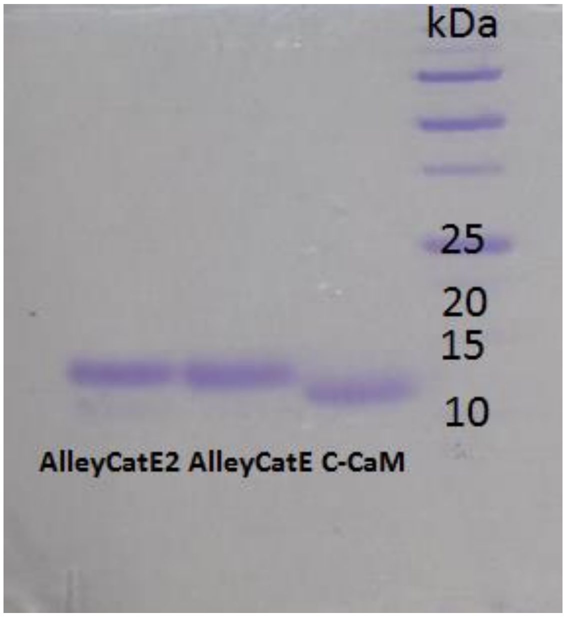


Figure S01. 10% SDS-PAGE of cCaM, AlleyCatE, and AlleyCatE2 proteins in 1X MES buffer.

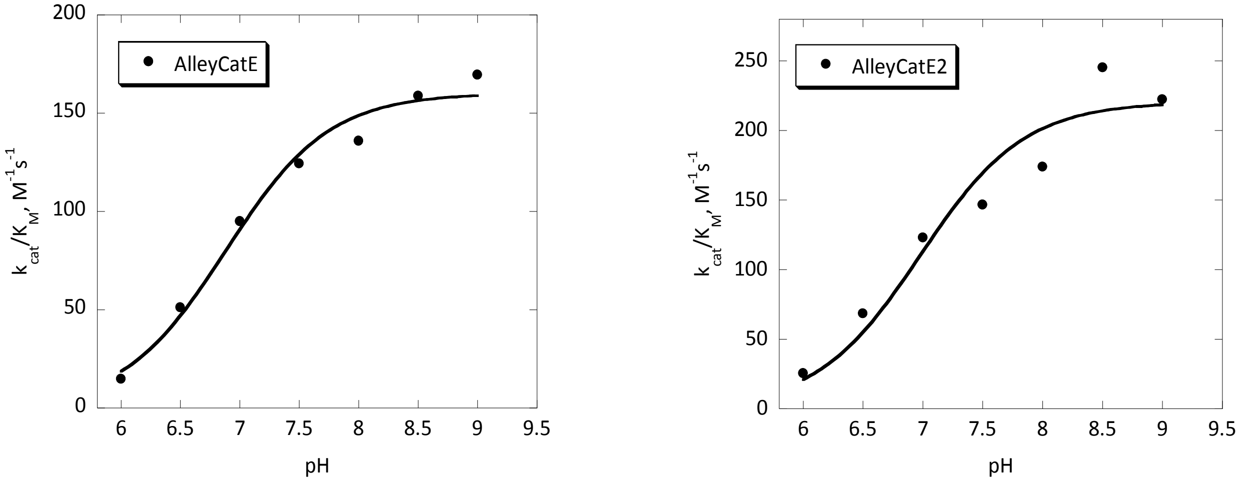


Figure S02. pH profiles of burst phase rate of AlleyCatE and AlleyCatE2. AlleyCatE shows a pK_a of 6.8 ± 0.07 with a k_{cat}/K_{Mmax} of 160± 5.2, AlleyCatE2 pK_a of 6.9 ± 0.14 with a k_{cat}/K_{Mmax} of 220± 14. Conditions: pH 4.5-9.5; 20 µM of substrate; 2.0 µM proteins. * pH's 4.5-5.5 were excluded. (Unpublished, TD2027 and TD20302)

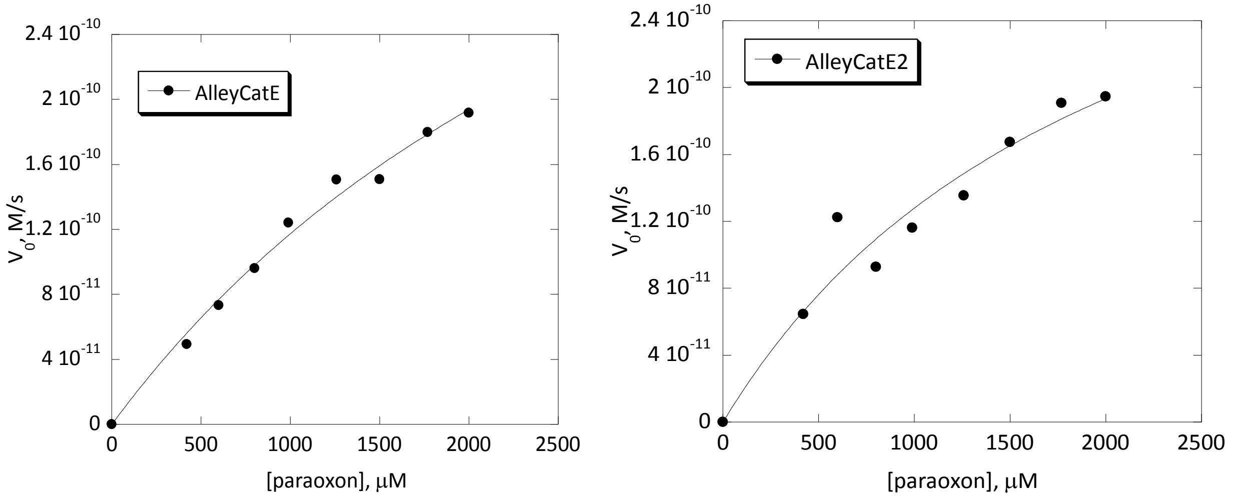


Figure S.03 Michaelis-Menten plot for hydrolysis of paraoxon catalyzed by AlleyCatE (left) and AlleyCatE2 (right) variant Conditions: 20 mM HEPES (pH 7.5); 100 mM NaCl; 10 mM $CaCl_2$; 10.0 μM proteins. k_{cat}/K_M for proteins are 0.10 and 0.16 ($M^{-1}s^{-1}$) respectively. (Unpublished, TD3019 and TD30211)

Saturation mutagenesis was tested on additional positions 92NNK, 109NNK, 124NNK, 127NNK, 145NNK, and 148NNK. From the screen, mutants M109P and M109W were obtained and pNPA hydrolysis was done on both. * M109P did not react with pNPA

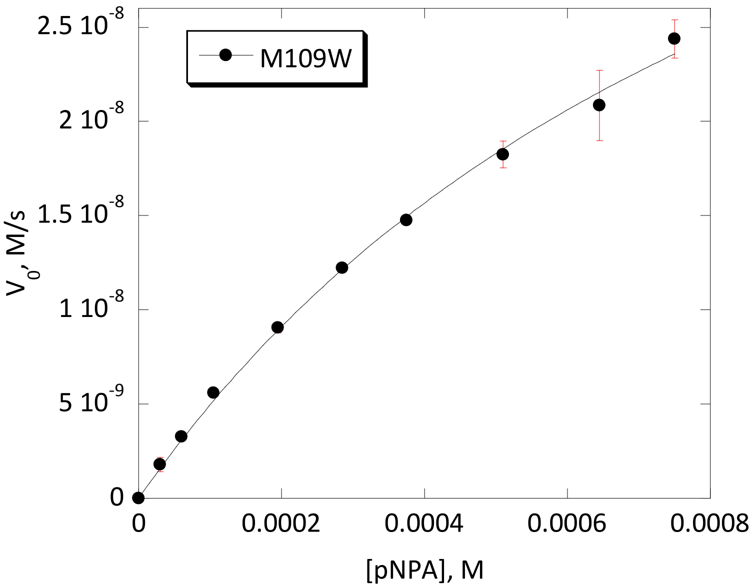


Figure S.04 Michaelis-Menten plot for hydrolysis of pNPA catalyzed by M109W AlleyCatE variant. k_{cat}/K_M of 9 ± 0.5 $M^{-1}s^{-1}$. (Unpublished, TD21071).

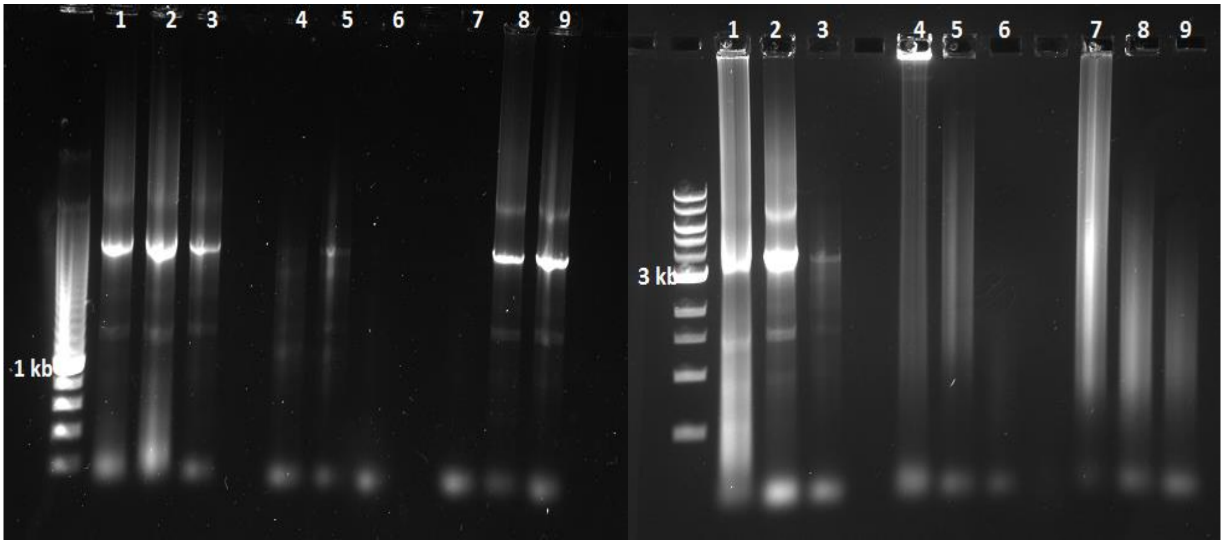


Figure S.05 Left: PCR products on 0.8% agarose gel on positions 92NNK (lanes 1-3), 109NNK (lanes 4-6), 124NNK (lanes 7-9). Right: 127NNK (lanes 1-3), 145NNK (lanes 4-6), and 148NNK (lanes 7-9). (Unpublished, TD20692 and TD2071).

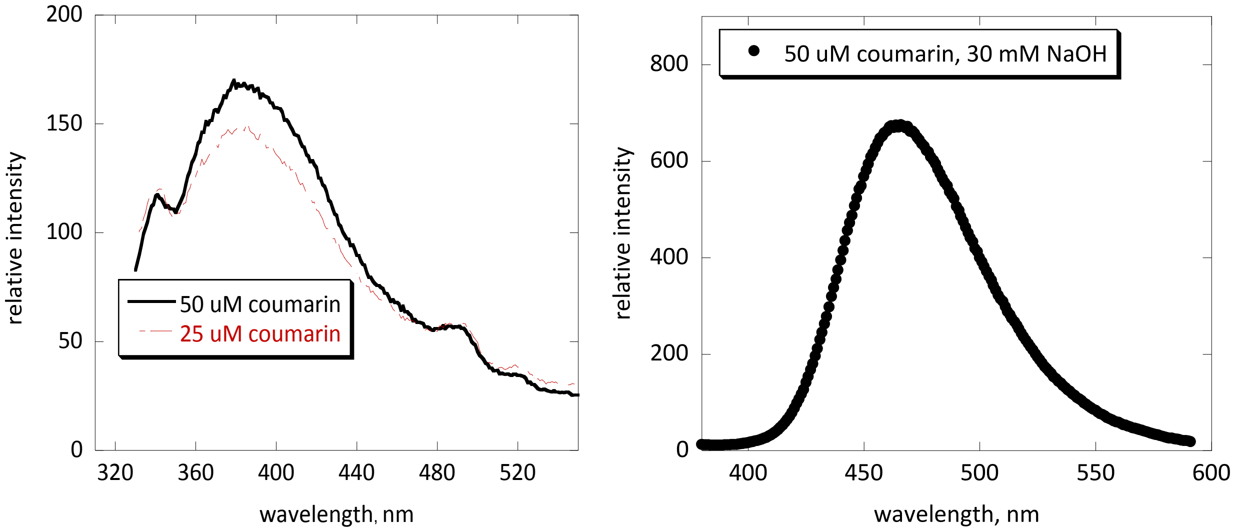


Figure S.06 Left: 50 μM and 25 μM coumarin no base emission scan excited at 310 nm with 10 nm slit. Substrate emits weakly at 383 nm. Right: 50 μM coumarin with 30 mM NaOH emission from 330-600 nm. (Unpublished, TD3036).

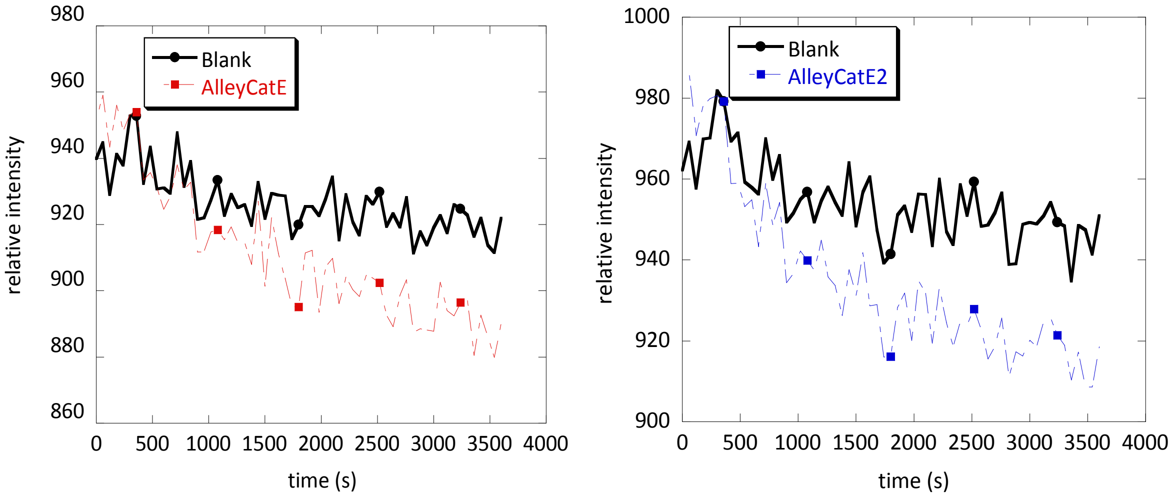


Figure S.07 Fluorescent platereader kinetic data of coumarin hydrolysis by AlleyCatE and AlleyCatE2 excited at 310 nm, emitted at 466nm for 1 hr. Final substrate concentration 750 μM, 10 μM protein concentration at pH 7.5. AlleyCatE or AlleyCatE2 does not hydrolyze coumarin. (Unpublished, TD3037).

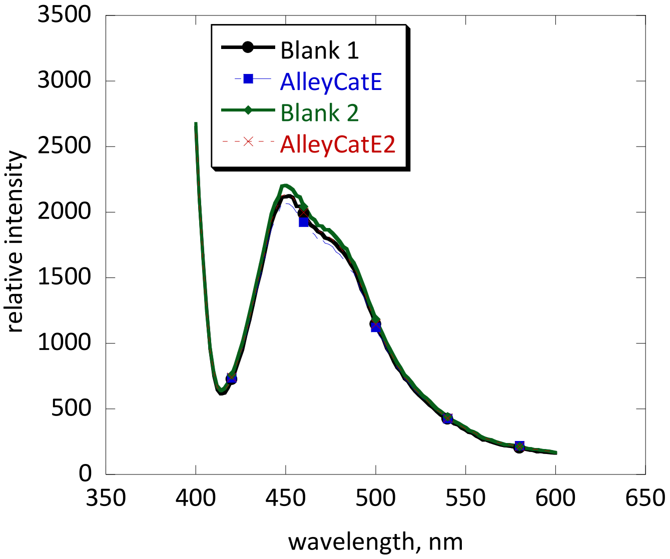


Figure S.08 Fluorescent plate reader emission spectrum at 750 μM coumarin substrate with 10 μM protein concentration pH 7.5 for 1 hr. Plate excited at 310 nm and emitted from 330-700 nm. (Unpublished, TD3037).

4. Small robust libraries in a thermostable enzyme for unnatural and natural reaction

4.1 Introduction

Mutating enzymes for the catalysis of different reactions can yield catalysts with an enhanced desired activity. We can fine tune active sites by altering the functionality in the hydrophobic cavity to achieve efficient catalysis. We can alter an enzyme's function using direction evolution, a technique used for the generation beneficial mutants.[40, 91]

Two new proteins served as starting scaffolds for catalyzing Kemp elimination and ester hydrolysis. The KE07 derived Kemp eliminases and tHisF, a thermostable enzyme from *Thermotoga maritima* [76, 92] were the protein targets for introducing new mutations. The target reaction for KE07 eliminase was Kemp elimination, which was previously discussed in details in Chapter 1. Many enzymes are deemed promiscuous in that they have the ability to catalyze a wide range of chemical transformations for unnatural and natural substrates. This feature has been linked to the native active site of all enzymes, but recently it has been discovered that enzyme catalysis can occur elsewhere.[93-94]

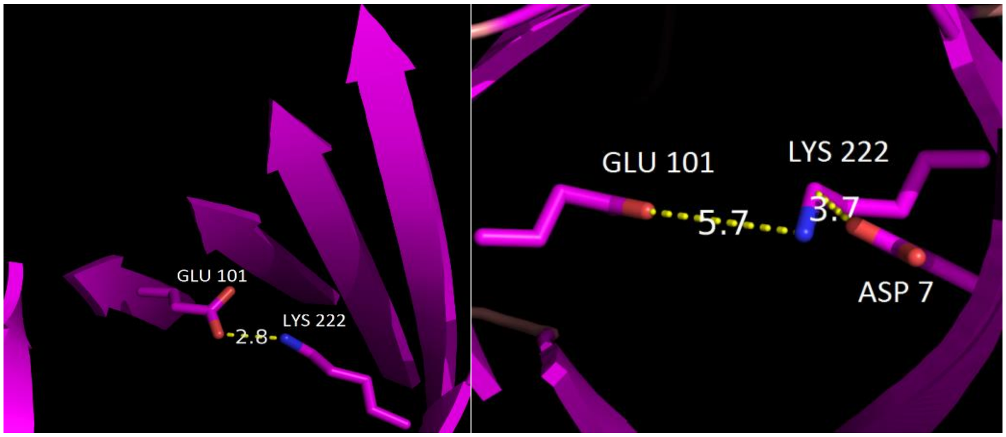
We are interested in discovering the origins of esterase activity in *tHisF* from *Thermatoga maritima* observed by Reetz *et al*.[93] Identifying catalytic residues apart from those in the active site will aim in the understanding of promiscuous enzyme behavior. Since tHisF primarily function in acid/base catalysis, we investigated its ability to function as an esterase in hydrolyzing para nitrophenyl butyrate (*p*NPB). In combination, we are examining both Kemp elimination and ester hydrolysis for two structurally similar designed enzymes.

4.2 Library Design and Characterization of KE07 eliminase

4.2.1 Structural features for identifying important active site residues

KE07 has the TIM barrel[95] like protein fold composed of 8 α-helices and a β barrel center.[96-98] The active site resides are located in the center of the beta barrel. An efficient enzyme must be able to stabilize transition state geometry so observing the key interactions of specific residues to perform catalysis was of great importance. As speculated by Baker *et al*, Lys222 was an important catalytic residue in that it was an acidic/H bond donor that stabilized the phenoxide transition state.[76] It also forms a hydrogen bond with the catalytic base Glu101 (Figure 4.1). Position Ile7 is situated at the bottom of the active site and in their previous designs were mutated to polar or charged amino acids. According to the crystal structure and their design structure, the distances between Lys222 and Glu101 is increased when the mutation Ile7Asp is introduced. This shows that Asp7 is effective at disturbing the salt bridge between Lys222-Glu101, making Glu101 a more reactive catalytic base for the reaction.

Baker *et al.* applied the theozyme approach, where they employed a carboxylate or a histidine carboxylate dyad motif to study key interactions in KE07 for catalyzing Kemp elimination (Figure 4.2). After designing the theozyme model using quantum mechanical calculations, RosettaMatch algorithm was applied where protein backbone positions were designed to match the predicted theozyme based model. Finally, the theozyme was grafted back onto the protein scaffold using RosettaDesign.[67]


GLU 101
2.8
LYS 222
GLU 101
LYS 222
5.7
3.7
ASP 7

A

B

Figure 4.1 (a) In the original crystal structure of KE07, the Glu101-Lys222 distance is 2.8 Å. (b) With the additional polar mutation D7 in close proximity, it disrupts the Glu101-Lys222 interaction and the distance increases, making Glu101 a more reactive base, PDB code: 1THF.

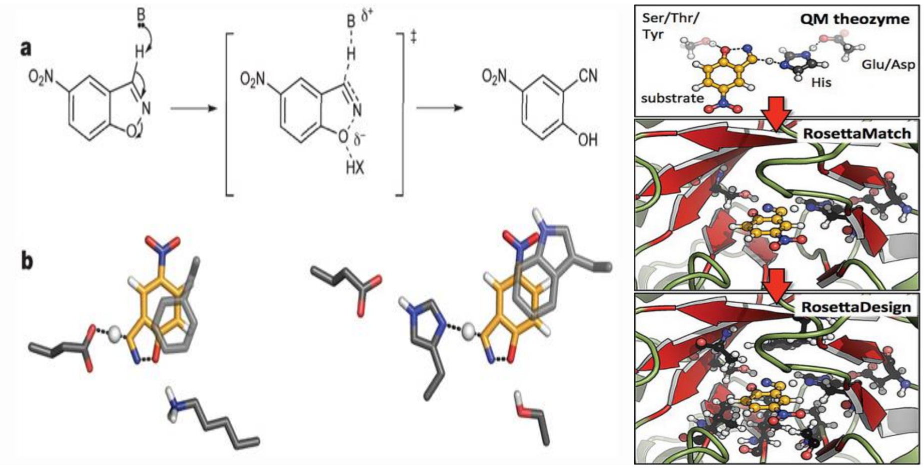

a
B
H
O₂N
N
O
B δ+
H
‡
O δ−
HX
CN
OH
b
Ser/Thr/
Tyr
QM theozyme
Glu/Asp
His
substrate
RosettaMatch
RosettaDesign

Figure 4.2 Theozyme model for Kemp elimination reaction. A: Kemp elimination displaying transition state. B: Designed carboxylate and histidine carboxylate motif incorporating surrounding active site residues; pi stacking residues (tryptophan) and H-bond donor lysine for stabilizing transition state. Right: Cartoon depiction demonstrating algorithms applied to theozyme model.[76]

4.2.2 Small robust PCR based libraries with a reduced amino acid alphabet

As previously mentioned, Baker *et al.* applied the theozyme approach, where they employed a carboxylate or a histidine carboxylate dyad motif to study key interactions in KE07 for catalyzing Kemp elimination.[67] However, we previously applied a minimalistic approach for catalyst design in calmodulin and now desire to see how different are these two approaches in terms of protein design? Can we take the evolved KE07 protein and examine the contributions from multiple residues? Amino acid Lys222 and Asp7 were closely studied and speculations were made based upon the possible influence from neighboring amino acids or how well it fits in the active site for catalysis. For catalyzing Kemp elimination reaction, we were interested in identifying and studying if there can be any possible beneficial interactions in the vicinity of the protein's active site to achieve catalysis. Considering these key residues, we are interested in making small robust libraries using a reduced amino acid alphabet via a megaprimer PCR method.[99] This method allows for generation of amplified gene and subsequent full vector amplification with target gene.

Three libraries were designed to test the importance of specific amino acids. Library A consisted of designing primers (see experimental) with NTT at position 128 and VTT at 130 on the sense strand and RSC at 171 (GSY for primer design) on the antisense strand. Having NTT covers all possible amino acid combinations that codes for L,V,F, or I amino acids with N representing equimolar amounts of each base. Having VTT codes for l, V, or I with V representing A,C, or G. Finally, having RSC can code for T,S,A, or G amino acids at position 171. Library B consisted of primers (see experimental) with RAT at position 7 and MTG at position 222 (CAK for primer design) on the antisense strand. The codon RAT codes for amino acids D or N, while MTG codes M or L. Library A is a substrate binding library where we are interested in introducing more hydrophobic residues in place of the native amino acids in those positions (Figure 4.3) to observe changes in substrate binding K_M and overall higher catalysis. Tyrosine is not hydrophobic, but polar so mutating this position and screening catalytic activity at this site will inform us if it is required for the reaction. Library B is an H-bond library. Since Lys222 is speculated to be an H bond donor, mutating this to similar sized nonpolar residues such as L or M we can investigate the true stability of Lys222 in stabilizes the transition state (Figure 4.2). Finally Library C, consisted of designing primers (see experimental) with KNT at position 9, VTS at 129, and DMT at 130. The forward primer contained position 9 KNT and the reverse primer contained AKH and SAB (reverse complement to DMT and VTS). Having DMT codes for amino acids S,A,T,Y,N, and D. Degenerate codon VTS codes for hydrophobic amino acids M,L,I, and V. Codon KNT is a much larger library coding amino acids F,S,Y,C,V,A,D, and G.

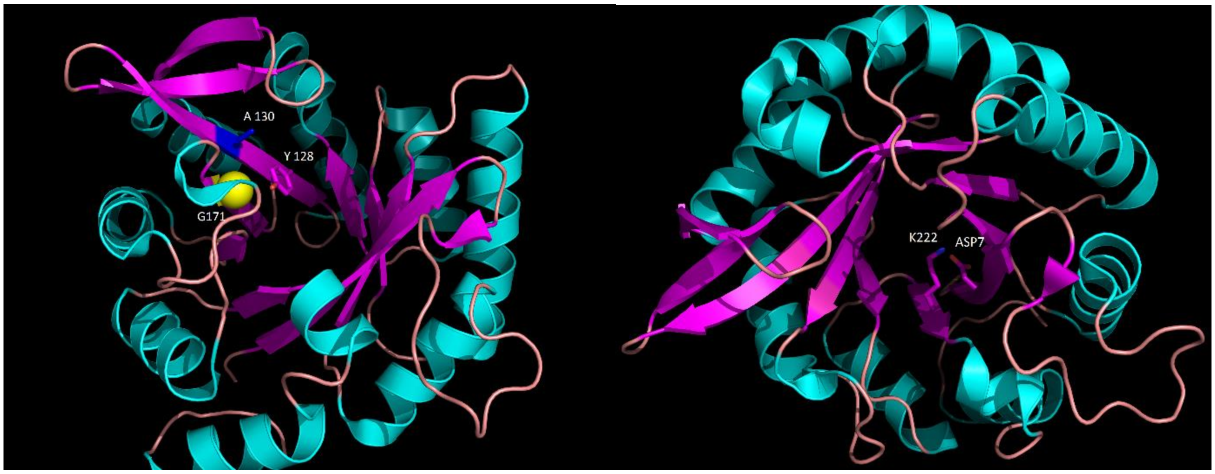


Figure 4.3 (left) Secondary structures α helices (cyan) and β sheets (magenta) of Library A with positions to be mutated represented by sticks A130 and Y128 and a sphere G171. (right) Secondary structures α helices (cyan) and β sheets (magenta) of Library B with positions to be mutated represented by sticks K222 and D7. PBD: 3iiv

4.2.3 Crude Screening for the Identification of active hits

Testing initial catalytic activity in crude cells is essential to ensure that the reaction has taken place and to identify the rates of active variants that are well above the wild type enzyme. Potential hits with higher activity were characterized by Sanger DNA sequencing. To ensure great sequence diversity at the desired position, the entire PCR mixture is also pooled and sampled together to observe library diversity amongst the degenerate codons (Figure 4.4). See appendix for libraries B and C diversities.Several plates for each position were then screened and confirmed in triplicate after identifying potential variants.

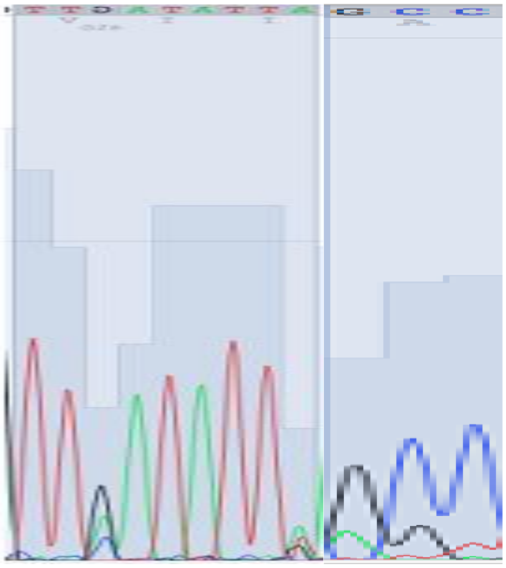


Figure 4.4 Sanger DNA sequencing of Library A diversity at positions Y128 (ATT), A130 (GTT), and G172 (GCC).

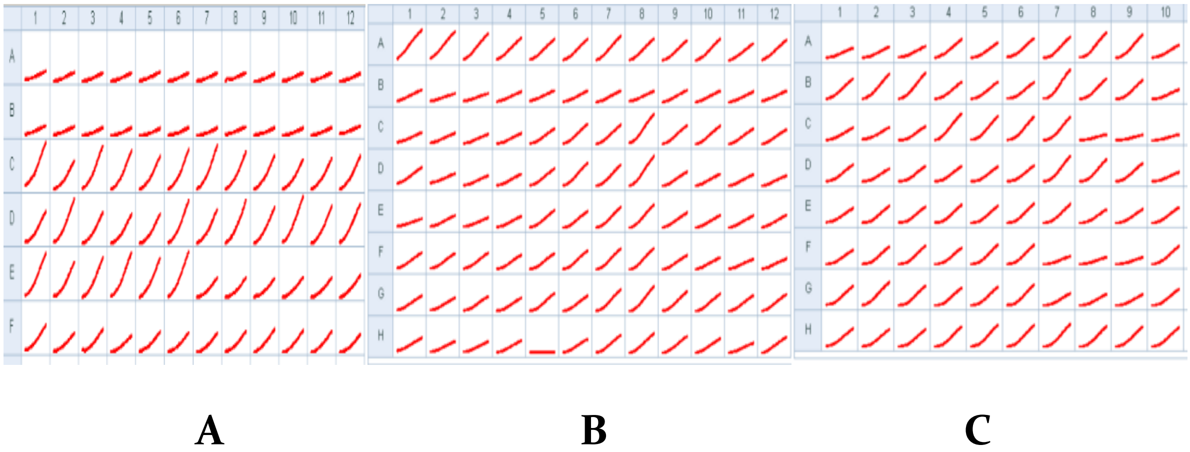


Figure 4.5 (a) Confirmation screen of combined library A and B hits with rows A-B blank, C1-C6= WT, C7-C12 = libA , D1-D6 = libA , D7-D1= G6, E1-E6= libA, E7-E12= libB, F1-F6= libB, F7-F12 libB (b,c) Confirmation screen of library C. (b) WT = G9-G12, others are confirmed variants from intial screens. (c) WT = C4-C7, others are confirmed variants from initial screen.

4.2.4 Kinetic Characterization of KE07 variants

From the initial screenings for libraries A, B, and C, eight variants were obtained and characterized for Kemp elimination reaction. Variants G171A and G171S were obtained for library A, K222M for library B, and I129V, A130S, A130T, I129VA130S, and A9G for library C. For the substrate binding library A, only position 171 became mutated after several rounds of screening. The tyrosine at position 128 did not become mutated which indicates its importance in the amino acid sequence. For library B, the lysine was speculated to be an H-bond donor to stabilize the transition state of the reaction, however it was mutated to methionine and the resulted variant was K222M. Catalytic activity was first tested for wild-type KE07 10/11G protein, G171S, and K22M. As shown in Figure 4.6, K222M maintains catalytic activity slightly higher than wild type enzyme. K22M has a catalytic efficiency of 830 $M^{-1}s^{-1}$ compared to KE07 10/11G of 793 $M^{-1}s^{-1}$. This result demonstrates that the lysine at position 222 is not a necessary H-bond donor for the reaction. Introducing a hydrophobic amino acid methionine at the same position catalyzes the reaction on par or slightly better than wild type. However, mutant G171S has a catalytic efficiency of 463 $M^{-1}s^{-1}$ much lower than KE07 10/11G and K222M. Variant G171A displayed a catalytic activity of 743 $M^{-1}s^{-1,}$ very much on par with wild type (Figure 4.7)

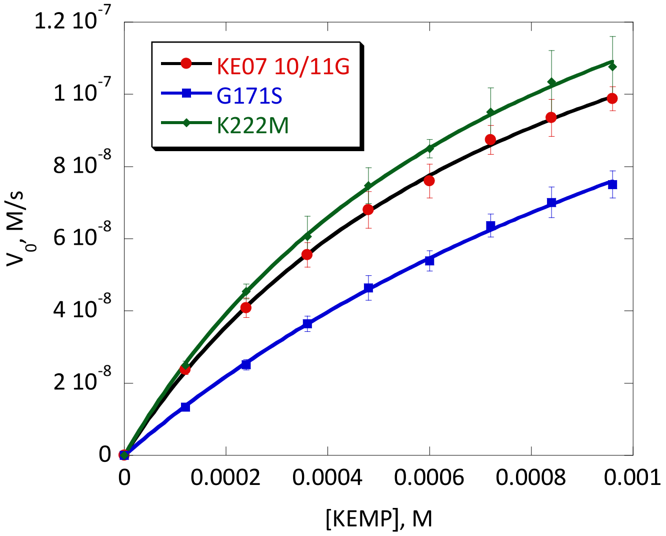


Figure 4.6 Kemp Elimination activity of KE07 10/11G, G171S, and K222M. Conditions: 25 mM HEPES (pH 7.0); 100 mM NaCl, 0.25 μM protein. Data were fit to Michaelis-Menten linear regime. (Unpublished, TD51251, TD51252, and TD51261)

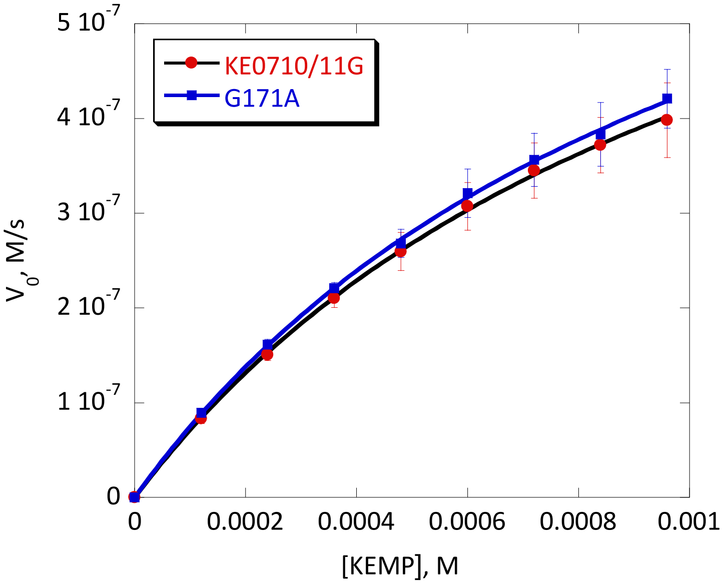


Figure 4.7 Kemp Elimination activity of KE07 10/11G and G171A. Conditions: 25 mM HEPES (pH 7.0); 100 mM NaCl, 1.0 μM protein. Data were fit to Michaelis-Menten linear regime. (Unpublished, TD5113 and TD5114).

Variants for Library C were next

characterized for catalyzing Kemp elimination. Both variants A130S and I129V both displayed a lower catalytic rate than KE07 10/11G. Catalytic efficiencies for A130S and I129V were 658 $M^{-1}s^{-1}$ and 720 $M^{-1}s^{-1}$ respectively (Figure 4.8). These results demonstrate that the protein cannot accommodate serine or valine at positions 129 or 130 with a greater activity than wild type protein. Also, mutants A9 and A130T showed an even lower activity than KE07 10/11 G of 558 $M^{-1}s^{-1}$ and 543 $M^{-1}s^{-1}$ respectively (Figure 4.9). Double variant I129VA130S displayed the highest catalytic activity compared to KE07 10/11G with a k_{cat}/K_M of 990 $M^{-1}s^{-1}$ (Figure 4.10). Comparing this double variant to the individual variants I129V and A130S, I129VA130S resulted in a lower K_M which explains the rise in catalytic efficiency.

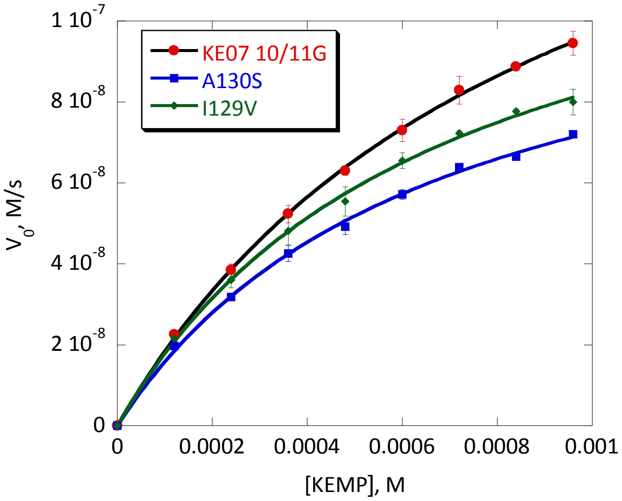


Figure 4.8 Kemp Elimination activity of KE07 10/11G, A130S, and I129V. Conditions: 25 mM HEPES (pH 7.0); 100 mM NaCl, 0.25 μM protein. Data were fit to Michaelis-Menten linear regime. (Unpublished, TD6081).

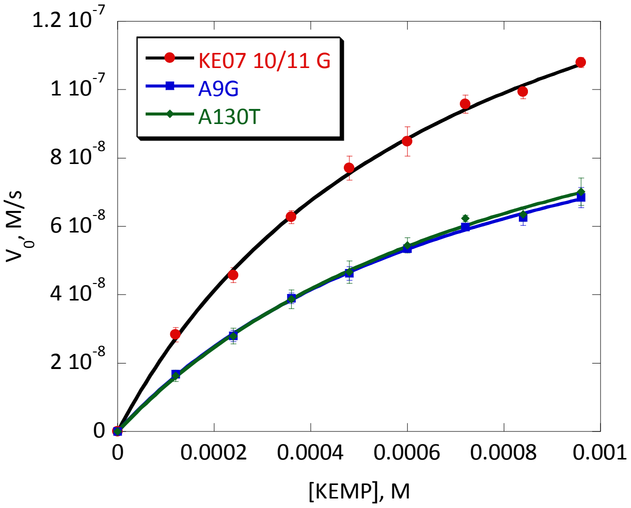


Figure 4.9 Kemp Elimination activity of KE07 10/11G, A9G, and A130T. Conditions: 25 mM HEPES (pH 7.0); 100 mM NaCl, 0.25 μM protein. Data were fit to Michaelis-Menten linear regime. (Unpublished, TD6126).

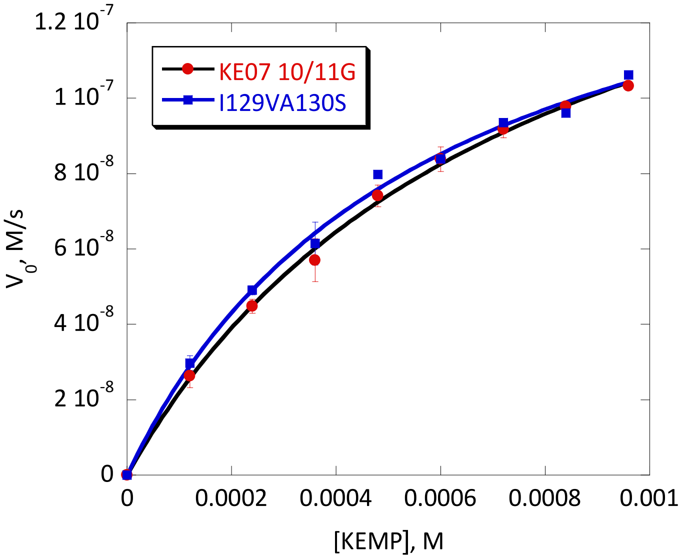


Figure 4.10 Kemp Elimination activity of KE07 10/11G and I129VA130S. Conditions: 25 mM HEPES (pH 7.0); 100 mM NaCl, 0.25 μM protein. Data were fit to Michaelis-Menten linear regime. (Unpublished, TD6127).

4.3 Exploring tHisF for esterase activity

4.3.1 Structural insight for exploring enzyme promiscuity

Catalytic promiscuity[100-101] is a unique property of enzymes. This property provides novel tools for organic synthesis of new compounds and broadened the application of enzymes.[102] Thermostable enzyme tHisF has been identified as a promiscuous enzyme, involved in the biosynthesis of histidine. *E. coli* contains the genes that encode for the proteins in the histidine biosynthetic pathway.[92, 103] The biosynthesis of histidine consist of a cascade of enzymatic events regulated by the products of genes HisA-HisI. HisF is an important component of a heterodimeric complex with HisH. Both subunits have unique roles in the synthesis of histidine. Shown in Figure 4.11 is the natural reaction of the complex where tHisH catalyze the hydrolysis of glutamine by producing ammonia and N'-[(5'-ribulosyl)-formimino]-5-aminoimidazole-4-carboxamide-ribonucleotide (5'-PRFAR). In the second step, tHisF uses ammonia in a condensation reaction with (5'-PRFAR) to give imidazole glycerol phosphate (IGP) and 5-aminoimidazole-4-carboxamide-1-β-D-ribofuranosyl 5'-monophosphate (AICAR).[103]

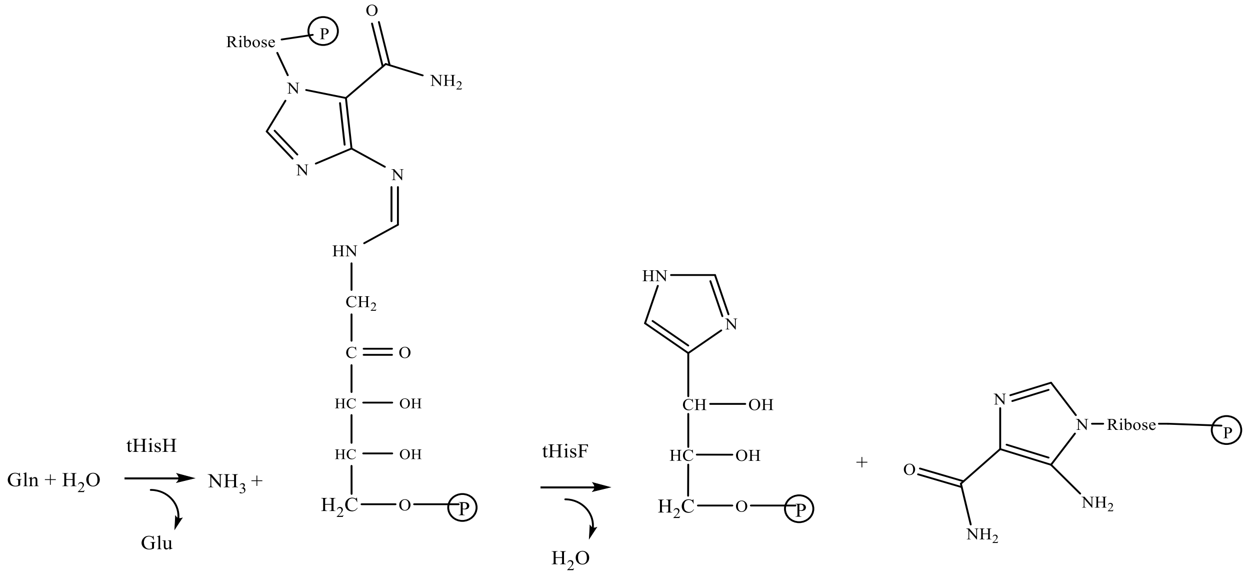

Gln + H_2O
tHisH
Glu
NH_3 +
Ribose
P
O
NH_2
N
N
N
HN
CH_2
C
O
HC
OH
HC
OH
H_2C
O
P
tHisF
H_2O
HN
N
CH
OH
HC
OH
H_2C
O
P
+
N
N
Ribose
P
O
NH_2
NH_2

Figure 4.11 Natural reaction of bienzyme complex tHisH-tHisF

The structure of tHisF is shown in Figure 4.12. It has a stable TIM barrel fold and is 27.7 kDa in size. Based on the structure, the active site is located at the center of the barrel highlighted by two phosphate ions. Reetz *et al.* studied the active site residues (D11, D130) and due to being involved in the natural function of the protein, he speculated that the protein can assume esterase activity at the same sites. However, applying the same mechanistic thinking with KE07, can there be any interactions for catalyzing *p*-nitrophenyl butyrate (pNPB) hydrolysis? Can we define the quantitative limit for esterase activity? We are interested in identifying an alternate catalytic site in this protein; therefore the two histidine at positions 84 and 228 away from the active site were mutated to glutamines and alanine. Glutamine is similar in size to histidine and alanine should be a complete knockout of function. The role of histidine in catalysis was thus explored.

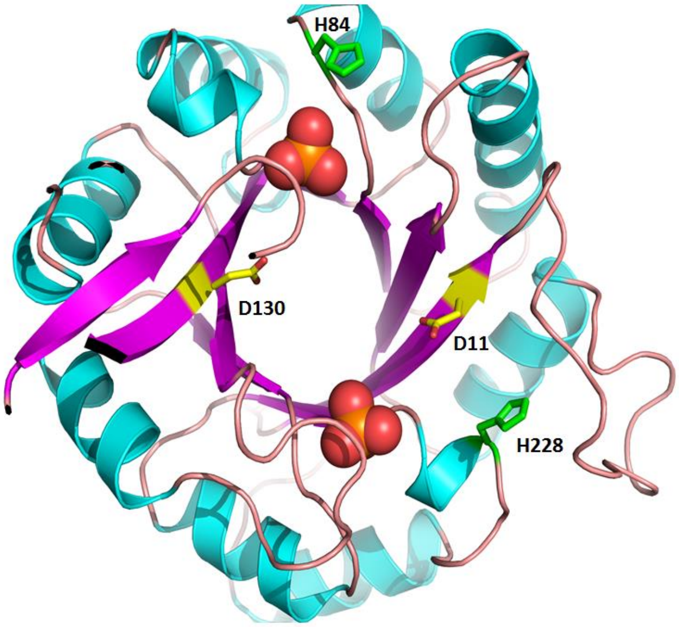


Figure 4.12 Crystal structure of tHisF (PDF code: 1THF) with α helices (cyan) and β sheets (magenta) and buffer phosphate as spheres. Asp11 and Asp130 are important catalytic residues. Histidine (green) were mutated to double glutamines or alanines.

4.3.2 Kinetic characterization pNPB hydrolysis

Variants H84QH228Q (QQ), H84AH228A (AA), and H228A (228A) were obtained by site-directed mutagenesis; the proteins were expressed and purified for catalyzing ester hydrolysis of *p*-nitro butyrate. Initially, UV spectroscopy was done to detect the solubility limit of pNPB (Figure 4.13). According to the results, 10% acetonitrile conditions worked best for the substrate and the solubility limit is 500 μM.

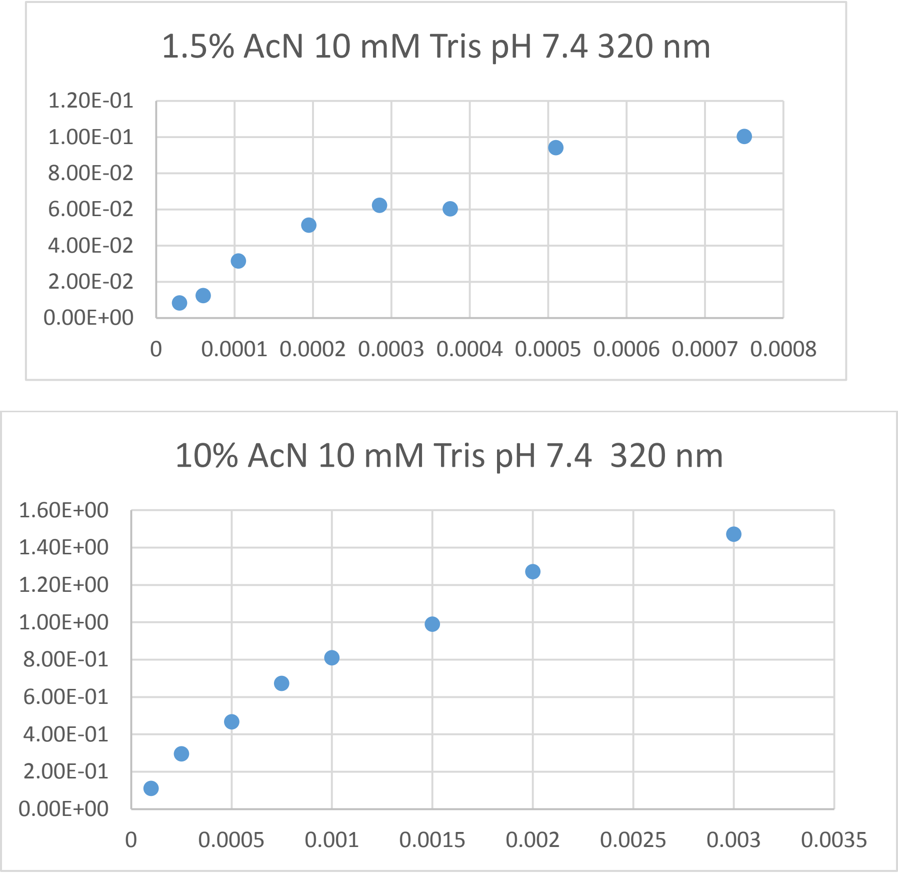


Figure 4.13 Solubility of pNPB at various substrate concentrations in the presence of 1.5% (top) and 10% (bottom) acetonitrile. (Unpublished, TD50571).

Kinetics were done in triplicate for all proteins in the presence of 10% acetonitrile and maximum 500 μM substrate concentration (Figures 4.14-4.17). According to the plots, we can observe the constant k_{cat}/K_M values for all variants compared to wild type tHisF protein which agrees well with the results of Reetz *et al*.

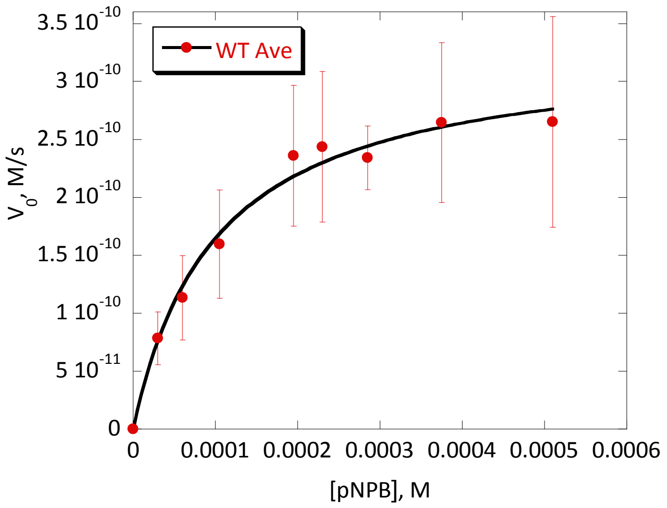


Figure 4.14 Michaelis-Menten plot of WT tHisF. The k_{cat}/K_M value obtained from linear fit (not shown) was 0.10 ± 0.02 at pH 7.4, 18 μM protein. (Unpublished, TD5060 and TD50621).

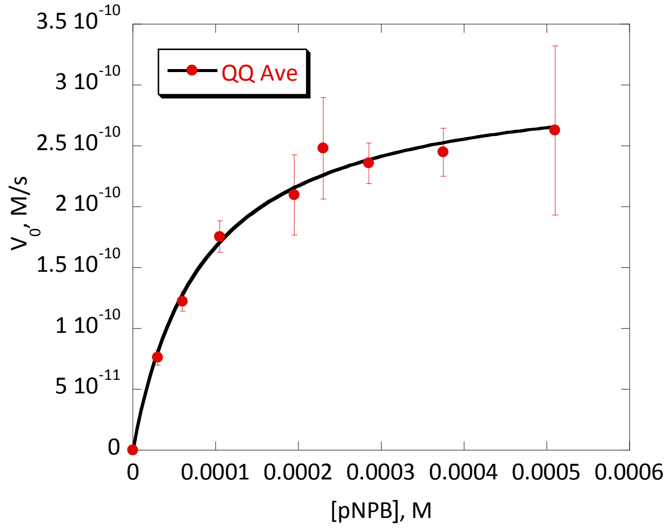


Figure 4.15 Michaelis-Menten plot of H84QH228Q (QQ) tHisF. The k_{cat}/K_M value obtained from linear fit (not shown) was 0.11 ± 0.01 at pH 7.4, 18 μM protein. (Unpublished, TD5060 and TD50621).

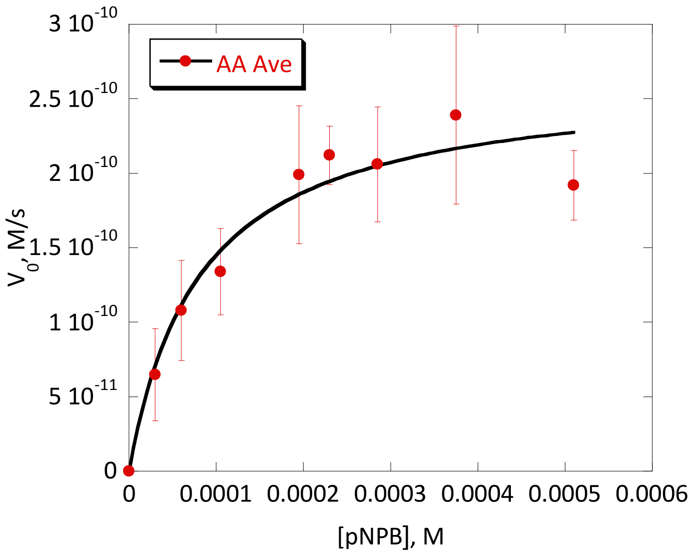


Figure 4.16 Michaelis-Menten plot of H84AH228A (AA) tHisF. The k_{cat}/K_M value obtained from linear fit (not shown) was 0.11 ± 0.01 at pH 7.4, 18 μM protein. (Unpublished TD50611 and TD50672)

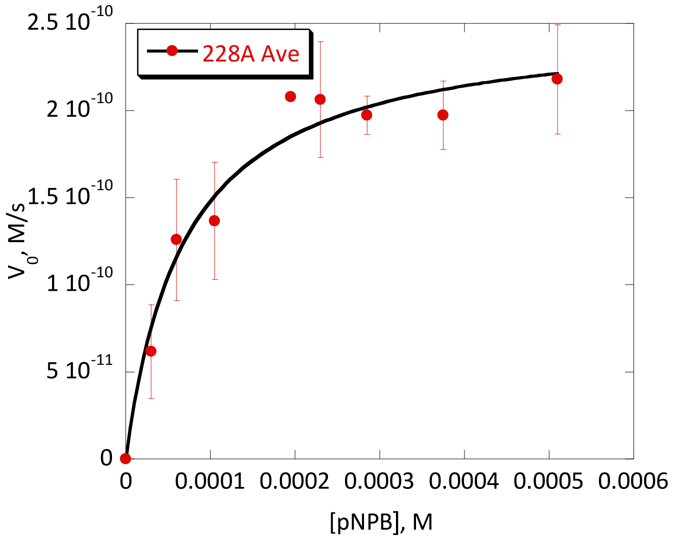


Figure 4.17 Michaelis-Menten plot of H228A (228A) tHisF. The k_{cat}/K_M value obtained from linear fit (not shown) was 0.11 ± .01 at pH 7.4, 18 μM protein. (Unpublished, TD50601 and TD50672).

As a control experiment, calcium binding protein calmodulin was utilized to catalyze the ester hydrolysis. Previously we have shown that calmodulin is an efficient esterase. [72] Although calmodulin and tHisF are two different proteins, we were interested in testing whether the reaction can be achieved with just simple hydrophobic sequestration of the substrate. We tested pNPB hydrolysis of calmodulin with lysines present and calmodulin without lysines[83] and compared those values with tHisF mutants (Figure 4.18). From the results, calmodulin with and without lysines does not enhance catalytic efficiency but is on par with tHisF mutants, demonstrating that the activity is not contributed by any residue but from hydrophobic association from the protein.

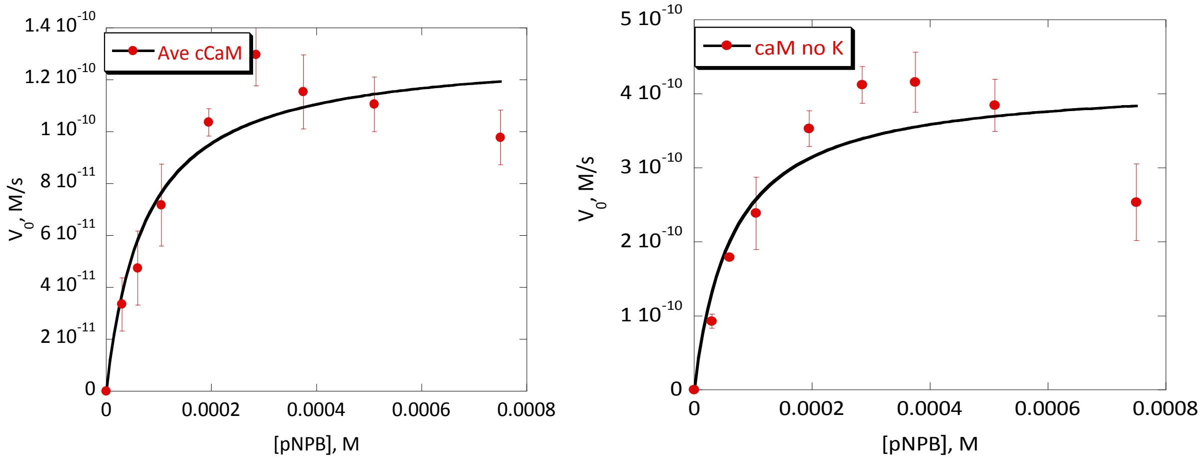


Figure 4.18 Michaelis-Menten plots of c-CaM and calmodulin no lysines (caM no K) with k_{cat}/K_M values of 0.13 ± 0.03 and 0.5 ± 0.1 respectively deduced from linear fit (not shown) at pH 7.4, 6 μM protein. (Unpublished, TD41481, TD41482, TD51551, and TD51552)..

4.3.3 Secondary structure analysis of tHisF variants

Circular Dichroism (CD) studies were done on all the proteins at 18 μM (Figure 4.19). CD data confirmed the secondary structural characteristics of the TIM barrel like proteins.

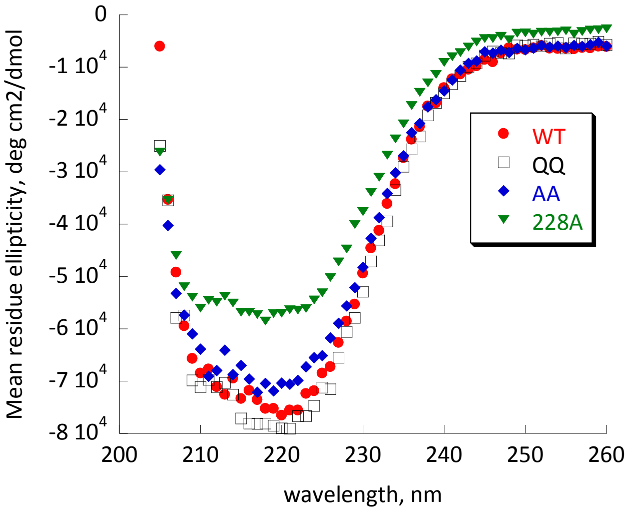


Figure 4.19 CD on all variants reveals the α-helical structure at pH 7.4 with 18 μM protein. (Unpublished TD50731).

. 4.4 Conclusion

Using a reduced degenerate codon design method, small robust libraries were designed with different functions in a given protein scaffold (KE07 10/11G) using a computationally inexpensive approach. Using site directed mutagenesis, specific mutations were introduced in the KE07 design and further evaluated as efficient kemp eliminases. This very stable scaffold was able to accommodate additional mutations for subsequent evolution. Libraries were created to elucidate the roles of K222 for hydrogen bond stabilization of the transition state and Y128. Results concluded that tyrosine is necessary at position 128 due to the protein being intolerant to mutations. Lysine at position 222 is not necessary for transition state stabilization as the K222M mutation produced a fully functional protein. High throughput screening allowed for quick hit identification in a matter of minutes. Enzyme promiscuity of tHisF was explored and showed that there can be a possible alternate catalytic site for ester hydrolysis reaction. With this robust technology of generating new functional catalysts, we are able to catalyze the Kemp elimination reaction and ester hydrolysis for different proteins.

Appendix Chapter 4 Supplemental Experiments

DNA gels from PCR

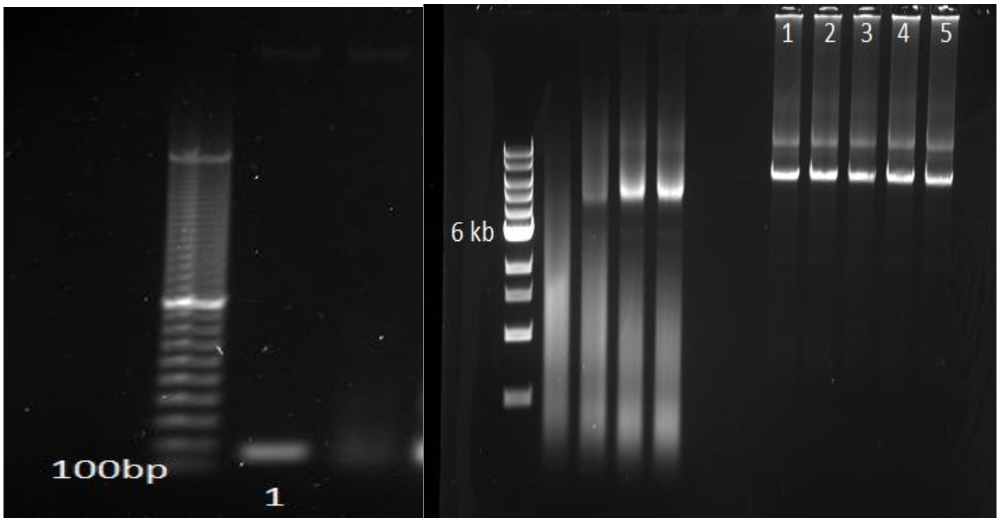


Figure S.01 Left: Gel A megaprimer product from library A. Right: Gel B PCR extension products generated from library A megaprimer at different annealing temperatures (Lanes 1-5) on a 1% agarose gel. (Unpublished, TD0822 and TD50841).

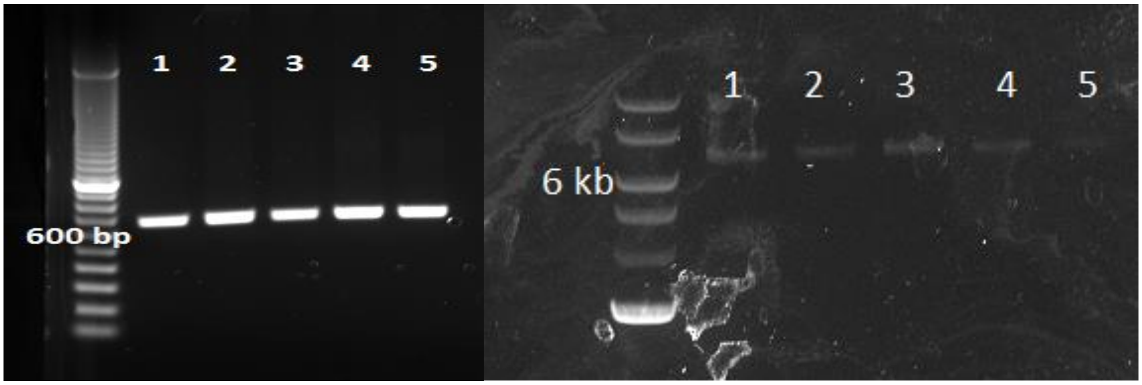


Figure S.02 Left: Gel A megaprimer product from library B. Right: Gel B PCR extension products generated from library B megaprimer at different annealing temperatures (Lanes 1-5) on a 1% agarose gel. (Unpublished, TD5086 and TD50891).

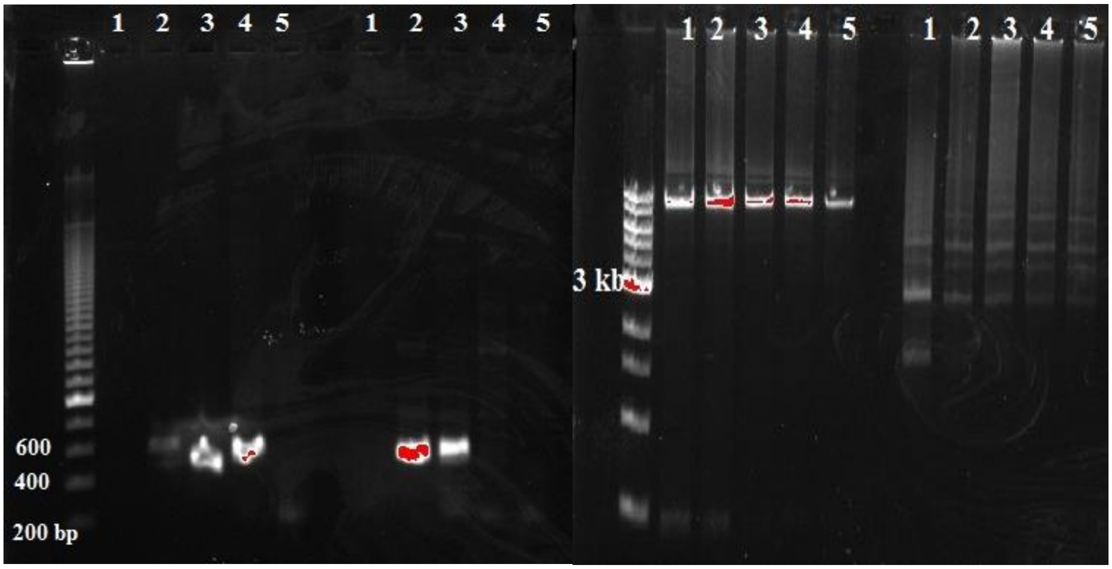

1 2 3 4 5 1 2 3 4 5
600
400
200 bp
1 2 3 4 5 1 2 3 4 5
3 kb

Figure S.03 Left: Gel of megaprimer product from library C. Right: Gel of PCR extension products generated from library C megaprimer at different annealing temperatures (Lanes 1-5) on a 1% agarose gel. Order: closest to ladder 1-5 (KE07 10/11G template), farthest 1-5 (K222M template). (Unpublished, TD51401 and TD5141)

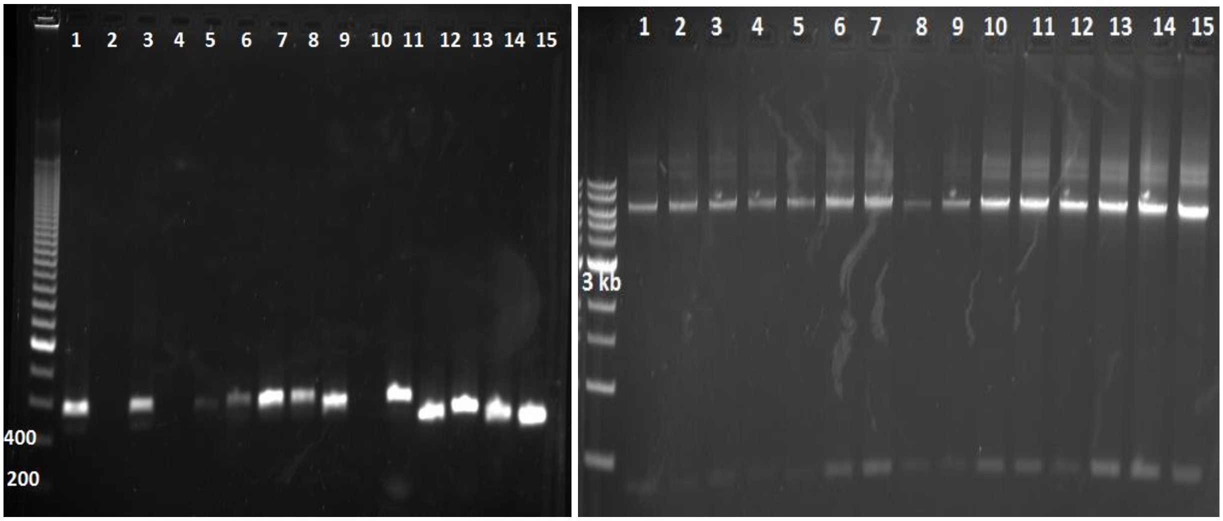

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
400
200
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
3 kb

Figure S.04 Left: PCR megaprimer for library C at different primer concentrations. Right: PCR extension using megaprimer 7, 11, and 13 from the left gel. (Unpublished, TD60891 and TD60922).

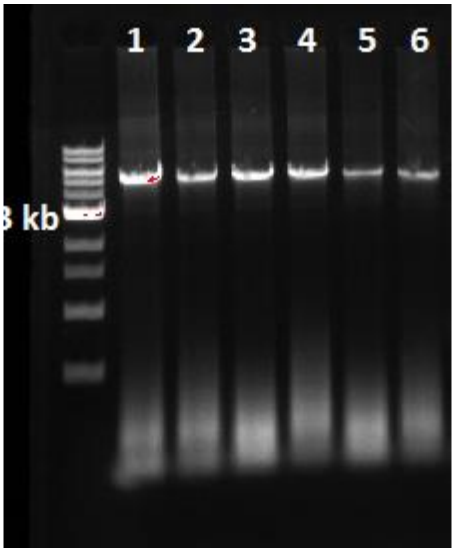


Figure S.05 PCR of D7NTT position on KE07 10/11G template. * K222M and I129VA130S templates were done but did not work. (Unpublished, TD61301).

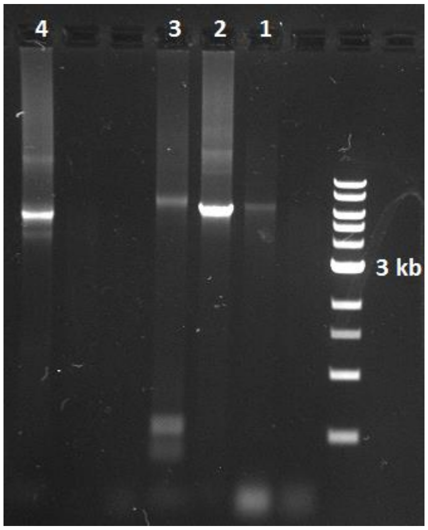


Figure S.06 0.8% agarose gel on PCR products. Shown in lane 1: H84AH228A (AA), lane 2: H84QH228Q (QQ), lane 3: H84QH228Q (QQ), lane 4: H84QH228Q (QQ). (Unpublished, TD4108).

Protein Gels

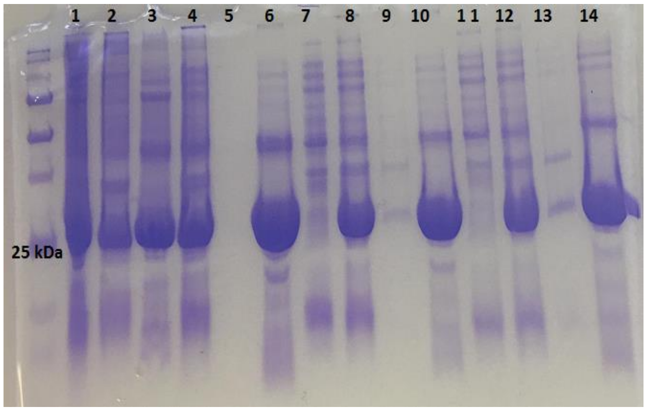

1
2
3
4
5
6
7
8
9
10
11
12
13
14
25 kDa

Figure S.07 15% SDS-PAGE of G172A and WT KE07 10/11 with different growth conditions (ZYM and IPTG 37C and 18C). Lane 6: soluble G172A ZYM 18C, Lane 10: soluble G172A IPTG 18C, Lane 14: soluble WT IPTG 18C. *other lanes are expression and debris pellet of the proteins. (Unpublished, TD51111).

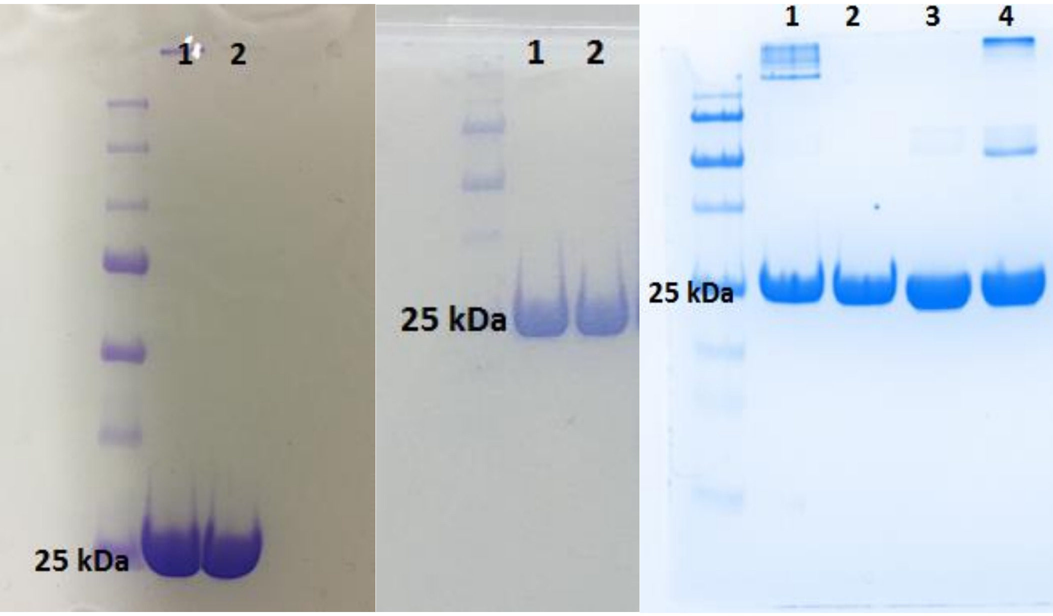

1
2
25 kDa
1
2
25 kDa
1
2
3
4
25 kDa

Figure S.08 15% SDS-PAGE: Left: 5 µg of G172A (lane 1) and WT (lane 2). Middle: I129V (lane 1), A130S (lane 2). Right: (lane 1-4) I129VA130S, A130T, WT, A9G. (Unpublished, TD61161, TD60791, and TD6125).

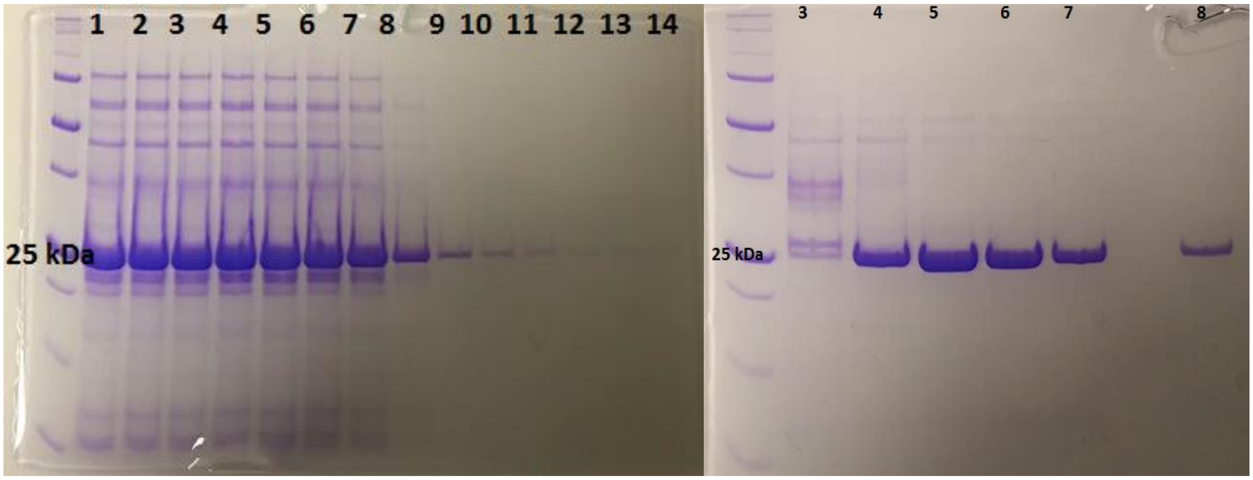

1 2 3 4 5 6 7 8 9 10 11 12 13 14
25 kDa
3 4 5 6 7 8
25 kDa

Figure S.09 Left: Fractions from G25 Desalting of tHisF. Right: Pure tHisF fractions collected after Q Sepharose purification. (Unpublished, TD41322 and TD1321).

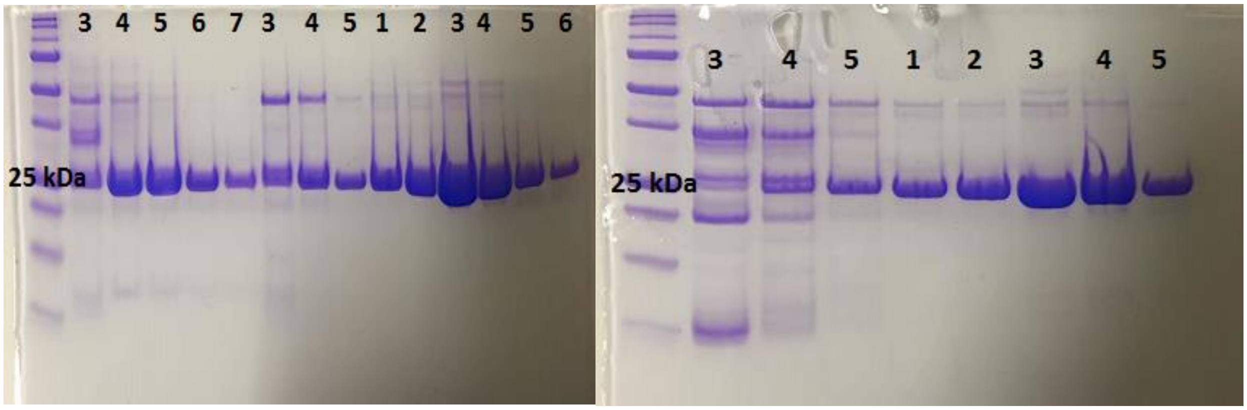

3 4 5 6 7 3 4 5 1 2 3 4 5 6
25 kDa
3 4 5 1 2 3 4 5
25 kDa

Figure S.010 Left: Fractions from Q Speharose of tHisF (3-7), QQ (lanes 3-5 at 200 mM NaCl) (lanes 1-6 at 250 mM NaCl). Right: Fractions from Q Sepharose of AA (lanes 3-5 at 200 mM NaCl, lanes 1-5 at 250 mM NaCl). (Unpublished, TD41341 and TD41351).

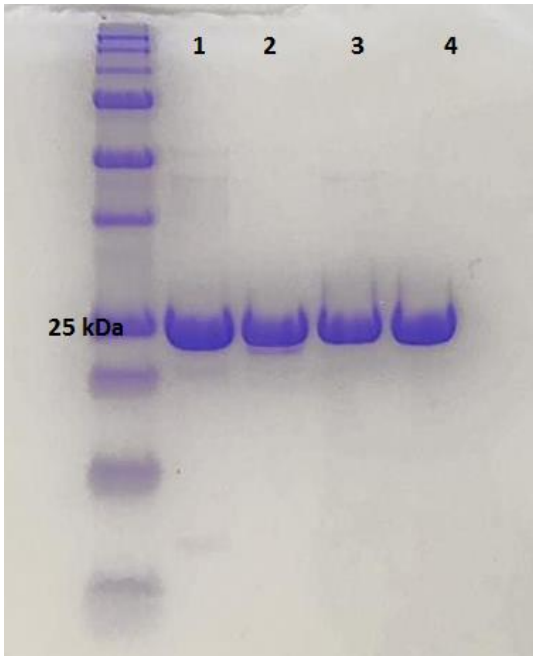


Figure S.011 Pure proteins buffer exchanged pH 7.4. Order from ladder: tHisF, QQ, AA, and 228A. (Unpublished, TD41381).

5. Experimental Procedures

5.1 Materials

All materials were purchased from the following suppliers and used without further purification: ChemImpex International (Wood Dale, IL), Pharmco-Aaper (Brookfield, CT), VWR (Radnor, PA), BioBasic. Inc (Amherst, NY), BioRad Laboratories Inc. (Berkeley, CA), or Sigma Aldrich (St. Louis, MO). BL21 (DE3) cells other E. *coli* cell lines were purchased from Agilent Technologies (Santa Clara, CA). Reagents and buffers were purchased from BioBasic, Inc.,and Santa Cruz Biotechnology, Inc. DNA oligonucleotides were purchased from Integrated DNA Technologies. Enzymes for cloning were purchased from Promega (Madison, WI) and New England Biolabs (Ipswich, MA). TEV (tobacco etch virus) protease was expressed and purified according to literature protocol. Lab supplies was purchased from Santa Cruz Biotechnologies (Santa Cruz, CA), World Wide Medical Supply (Bristol, PA), and/or VWR (Radnor, PA). Racemic *p*-nitrophenyl-(2-phenyl)-propanoate was prepared from 2-phenylpropionic acid and 4-nitrophenol using previously reported procedure.[84] The enantiomerically pure (R) - and (S)-p-nitrophenyl-(2-phenyl)-propanoates were prepared from the corresponding commercially available enantiomers of 2-phenyl-propionic acid with the same protocol.

5.1.2 Protocol Preparation

All medias, protein gels, and PCR procedures were prepared using a standard protocol according in the tables below. Some alterations with PCR such as temperature, annealing time, and extension time were done if needed.

ZYM5052 media is prepared as followed:

Component	Final Concentration in ZYM (mM)	Molecular Weight (g/mol)	Mass/Volume
N-Z Amine	1%		10 g
Yeast Extract	.5%		5 g
Na2HPO4	25	268.07	6.7 g
KH2PO4	25	136.09	3.4 g
NH4Cl	50	53.49	2.67 g
Na2SO4	5	142.04	.71
Glycerol (40%)	.5%		12.5 mL
Trace metals	.2X		0.2 mL
H2O	Add to 1L		

Autoclave: Liq 30 or 40, lactose is added to a final 0.2% and dextrose to a final 0.05%.

M9 Labelled Media:

Component	Mass/Volume
Mg salts basis	100 mL
MgSO4 (1M)	1 mL
CaCl2 (1M)	100 µL
1000X trace metals (.5X)	0.25 mL
Glucose (dextrose)	2 g
100X MEM vitamins (10X)	5 mL
^{15}NH4Cl Shake well then sterilize	0.5 g

10%, 15% SDS-PAGE

All components were added and mixed. Isopropanol was added to the top to prevent bubbles. After solidification, the stacking gel was added and a comb was inserted to created wells.

Separating Gel

Component	Volume (mL)
30% Acrylamide	2.5
4X Tris pH 8.8	1.88
$H_2$0	3.13
10% APS	0.050
TEMED	0.010

Stacking Gel

Component	Volume (mL)
30% Acrylamide	0.325
4X Tris pH 6.8	0.625
$H_2$0	1.5
10% APS	.013
TEMED	.003

PCR Standard Protocols

Pfu

Step	Temp °C	Time	Cycles
Initial Denaturation	95	1-2 min	1
Denaturation	95	0.5-1 min	25-30
Annealing	Primer Tm-5°C	30s	↓
Extension	72-74	1 min per kb	↓
Final Extension	72-74	5 min	1
Hold	4	indefinite	1

Phusion

Step	Temp °C	Time	Cycles
Initial Denaturation	98	30 s	1
Denaturation	98	5-10 s	25-35

Annealing	Primer Tm+3°C	10-30 s	↓
Extension	72	15-30 s per kb	
Final Extension	72	5-10 min	1
Hold	4	indefinite	1

5.2 Computational Modeling

5.2.1 Kemp eliminase F92E Design

SLMKDTDSEEEIREAFRVFDKDGNGYISAAELRHVMTNLGEKLTDEEVDEMIREADIDGDGQVNYEEFVQMMTAK sequence was modeled onto the structure of C-terminal domain of CaM (PDB ID code 1QS7) by maintaining the backbone and side chains of the common amino acids and building the amino acids that varied by finding the lowest energy (CHARMM force field) conformation.[104]

AutoDock version 4.06 [105] was used to dock 5-nitrobenzisoxazole into CaM (maximum rotational/translational step sizes 50°, 2 Å; population size 150; 50,000 generations). The docking parameters were systematically varied in independent runs all achieving approximately the same docked pose.

First, 100 side chain conformations (rotamers) for the aspartate/glutamate position and 50 rotamers for any residue within 8 Å of the Asp/Glu were built using the Molecular Software Libraries (MSL, manuscript in preparation). Then, the pairwise energy table (CHARMM force field) was computed. The global minimum rotamer configuration was found using a cyclic simulated annealing Monte Carlo protocol, assuming the reference state energy to be similar. The final models were processed through a 100-step constrained CHARMM minimization using the Adopted Basis Newton–Raphson algorithm.

A superrotamer of glutamate-5-nitrobenzisoxazole was created based on QM calculations[76, 106]. The superrotamer models were generated using the same protocol as specified above with the number of rotamers doubled. Partial charges were derived for 5-nitrobenzisoxazole using AM1-BCC. The bonded energy parameters for 5-nitrobenzisoxazole not present in CHARMM, were replaced with the most chemically similar terms.

5.2.2 Esterases

A high resolution crystal structure of calmodulin (PDB entry 1CLL) [73] served as a starting point of the design. The side chain placement was computationally optimized with fixed backbone (fixbb protocol) using Rosetta.[31] Mutations in the resulting model were computationally introduced by changing the amino acid identity in positions A88, F92, L105, V108, M109, M124, A128, F141, M144, M145 to a histidine, and side chain placement was again computationally optimized with a fixed backbone (fixbb protocol) using Rosetta. The Rosetta score for the mutants was compared to that of the original structure and reported. The models of R and S isomers of p-nitrophenyl-(2-phenyl)-propanoate were created with MarvinSketch (ChemAxon) and optimized with UCSF Chimera [107] and ADT 1.5.6 .These models were then noncovalently docked into the computationally derived models of CaM mutants using Autodock Vina software.[87] At least 20 docking models for each mutant were generated. Then, the distances between the δ and ε atoms (both nitrogen and carbon atoms were considered to account for a possibility of a 180° rotation around the (Cβ–Cγ bond) in the histidine ring and the carbonyl carbon were computed. The poses for the substrates were considered capable of supporting catalysis only if (a) the distance between carbonyl carbon of the ester and the nitrogen atom of imidazole ring (either Nε or Nδ) was less than 3.5 Å and (b) the $O{=}C_{carbonyl}$-$N_{imidazole}$ angle was between 100° and 140°.

5.3 DNA Cloning/ Mutagenesis

All of the genes for full length CaM 7th R were cloned into PEXP5-NT/TOPO expression vector and c-terminal calmodulin (AlleyCatE and AlleyCatE2) were cloned into PMCSG49 vector by ligation independent cloning [108] Both vectors contain an N-terminal His_6-tag and a tobacco etch virus (TEV) protease recognition site (ENLYFQ/S).

Error prone PCR (ep-PCR) was done using GeneMorph II EZ clone Mutazyme kit (Stratagene) and saturation mutagenesis was done using *Pfu* or *Phusion* standard protocol. PCR products were confirmed by DNA electrophoresis. PCR products from ep-PCR were digested with restriction enzyme *DpnI* and transformed into BL21 PLysS cells. Colonies were picked, grown in LB, and expressed in ZYM5052 for crude cell screening. PCR products from saturation mutagenesis were digested with restriction enzyme *DpnI* and transformed into XL10 cells for Sanger DNA sequencing (Genscript or Genewiz Inc.). For NNK codon derived mutants, combinatorial PCR based libraries were generated and screened against wild type protein. To test for a higher frequency of mutations~800 colonies were screened.

5.3.1 Ligation Independent Cloning

S81R CaM 7thR and K94M CaM 7thR genes were amplified via PCR (50 μL volume) and gene was extracted using DNA gel extraction kit. T4 polymerase was performed on the gene via PCR reaction (20 μL volume). Primers were designed that contained 5' overhangs to properly anneal to vector overhangs. From the reaction, 2 μL of the gene was added to 0.73 μL vector and incubated at room temperature for 5 min, then 1 μL of 25 mM EDTA was added, incubated at room temperature for 5 min. Sample was transformed and sequenced.

5.3.2 Screening Procedure for Kemp elimination

All libraries are screened in a high-throughput fashion using a Multiskan spectrum platereader (Thermolab systems) or a Biotek (Biotek Instruments, Inc) plate reader. ZYM cultures were spun down at maximum speed for 7 minutes at 20 °C and lysed with Cell Lysis Buffer pH= 7.4 (100mM NaCl, 20mM HEPES, 10mM $CaCl_2$, 0.2% *Triton* x-100). Into 96 well plates, 150 μL Cell lysis Buffer was added with 50 μL of product using multichannel pipettes. Plates were screened for 30 readings every minute at 380 nm for CaM 7th R mutants and 15 readings every 40 seconds for KE07 10/11G mutants for Kemp elimination reaction. All hits were identified and confirmed in triplicate. DNA was isolated using a Biobasic kit, transformed into XL10 *E. coli* cells. Colonies were inoculated into LB and antibiotic (Ampicillin or Kanamycin) overnight at 37 °C. DNA of the mutated genes were sent for sequencing.

5.3.3 Crude cell Lysate Screening and Directed Evolution

To test the activity of various proteins to catalyze ester hydrolysis in crude cell lysates, the genes for c-CaM, AlleyCatE, CaM, and CaM M144H (full length protein) mutants were transformed into *E. coli* BL21 (DE3) pLysS cells. Individual colonies were cultured into 200 µL of LB containing 100 µg/mL of ampicillin in 96-well plates and grown for 6–7 h at 37 °C. A 10 µL portion of the resulting culture was used to inoculate ZYM-5052 autoinduction media containing 100 µg/mL of ampicillin and grown overnight at 37 °C. The following day, cells were harvested and lysed with cell lysis buffer containing 20 mM HEPES, 10 mM CaCl2, 100 mM NaCl, 0.2% Triton X-100, pH 8.0. Hydrolysis of 25 µM p-nitrophenyl-(2-phenyl)-propanoate (pNPP) by the crude cell lysates was monitored spectrophotometrically at 405 nm. For directed evolution studies, the screening was done with the full length CaM to improve expression levels. The presence of the N-terminal domain does not affect catalytic activity. Saturation mutagenesis to generate all possible single mutants in positions D80, L105, H107, V108, L112, A128, A88, F89, and V91 was done by following a standard mutagenesis protocol with Pfu Turbo DNA polymerase (Stratagene) using an NNK codon and AlleyCatE as a template. The resulting library of about 900 mutants was screened for hydrolytic activity, and the mutant that showed the best steady-state activity in the crude lysate (AlleyCatE2) was chosen for further characterization.

5.4 Protein Expression and Purification

All proteins were expressed by LB IPTG induction or ZYM5052 auto induction. All proteins were purified by nickel affinity chromatography and additionally by anion exchange (Q Sepharose), if needed. Detailed information for specific proteins are outlined below.

5.4.1 Esterases and Kemp eliminases

The vectors containing the genes of interest were transformed into *E. coli* BL21(DE3) pLysS cells and expressed using ZYM-5052 autoinduction medium. The cells were harvested by centrifugation and lysed by sonication in a buffer containing 25 mM TRIS, 20 mM imidazole, 10 mM CaCl2, and 300 mM NaCl (pH 8) on ice with a protease inhibitor (phenylmethylsulfonyl fluoride, PMSF) added. The crude cell lysate was then centrifuged at 17,000 g for 1 hour, and the supernatant was applied onto a Ni-NTA column (Clontech). After several rounds of washing, the protein was eluted with buffer containing 25 mM TRIS, 500 mM imidazole, 10 mM CaCl2, and 300 mM NaCl (pH 8). To remove imidazole the buffer was exchanged to 20 mM HEPES, 10 mM CaCl2, 100 mM NaCl (pH 7 for Kemp eliminases) and (pH 7.5 for esterases) by applying the purified sample onto a BioRad 10 DG desalting column. $\varepsilon = 4450\ M^{-1}cm^{-1}$ full length proteins, $2967\ M^{-1}cm^{-1}$ C-terminal proteins.

5.4.2 KE07 10/11G variants

Expression was done by inoculation of a single colony of the mutated gene into 4 mL LB and 4 µL kanamycin (kan) for 6-8 hours. After inoculation, cultures were transferred into 1 L LB+ 1 mL kan until OD_{600} of .6. Culture was induced with 0.5 mM IPTG at 18°C overnight and harvested by centrifugation. Pellets were kept on ice and 15 mL of Lysis wash buffer pH = 7.5 (25 mM HEPES, 100 mM NaCl) was added along with PMSF. Sonication was done for 10 min total with 30 second pulses. Lysed contents were centrifuged at 20,000g, 4°C, for 45 minutes. Ni-NTA columns were pre-equilibrated with lysis wash Buffer (25 mM HEPES, 100 mM NaCl) pH= 7.5 and washed with 20 mM imidazole. Proteins were eluted with buffer containing (25 mM HEPES, 100 mM NaCl, 250 mM imidazole) pH= 7.5. A UV-Vis Spectrometer was used to measure absorbance of protein and calculate concentration. Next, 3 mL of the purified proteins were then buffer exchanged into kinetic Buffer (25 mM HEPES, 100 mM NaCl) pH= 7 to further study kinetics. The protein extinction coefficient is 18,450 $M^{-1}cm^{-1}$.

5.4.3 tHisF variants

For all mutants expression was conducted with BL21(DE3)pLysS cells. A 100 mL overnight LB culture supplemented with ampicillin (100 mg/mL) and chloramphenicol (34 mg/mL) was inoculated overnight at 37°C and transferred into TB-medium with the same antibiotics. Overexpression was induced at an optical density at 600 nm (OD_{600}) between 0.4 and 0.6 by addition of IPTG to a final concentration of 1 mM. After another 2h, the cells were harvested by centrifugation (15 min, 6000 rpm, 4 °C), washed with 100 mM Tris-HCl pH 7.4 buffer and centrifuged again (25 min, 8000 rpm, 4 °C). The cell pellet was resuspended in 150 mL 10 mM Tris-HCl pH 7.4 buffer and supplemented with 20 µL of benzonase (Boehringer). After sonication, the cell-suspension was treated at 72 °C for 35 min (denaturation of non-thermostable proteins), centrifuged (14000 rpm, 45 min, 4 °C) and the clear supernatant was collected. SDS-Page (12.5%) showed single band purity for all mutants studied. tHisF mutants were purified by Q Sepharose anion exchange purification. Proteins were lysed in 10 mM Tris, washed with 100 mM Tris pH 7.4, sonicated and centrifuged. Proteins were desalted using a G25 Superfine medium resin before applied to Q sepharose. The desalting column was washed 3 times with MilliQ water then equilibrated 3 times with starting buffer 10 mM Tris. Supernatant was applied to the column and fractions were collected and BCA assayed for protein. G25 fractions containing protein were combined and applied to Q sepharose where the protein was eluted with 250 mM NaCl. Final buffer exchange was done in 10 mM Tris pH 7.4 buffer for kinetics.[93] The protein extinction coefficient is 11,460 $M^{-1}cm^{-1}$.

5.4.4 Expression and Purification of Isotopically Labeled Proteins

U-^{15}N,^{13}C-labeled AlleyCatE was expressed with double antibiotic resistance following the previously described procedure. The plasmid encoding the protein sequence with an N-terminal His6-tag followed by the TEV protease recognition site (ENLYFQ/S) was transformed into BL21(DE3) pLysS cells. The next day, a single colony was inoculated into a 50 mL LB and the resulting culture was grown overnight at 30 °C. The next morning the overnight culture was added to 1 L of 2× LB and incubated at 37 °C and 230 rpm until OD600 reached 0.6–0.7. The cells were harvested by centrifugation for 8 min at 1000 g and 30 °C, washed first with unlabeled M9 minimal medium and then with 25 mL of M9 minimal medium that contains $^{15}NH_4Cl$ and uniformly ^{13}C-labeled glucose. The cell pellets were resuspended in 300 mL of M9 minimal medium solution containing 15NH4Cl as a nitrogen source, and ^{13}C-labeled glucose as a carbon source, and incubated at 37 °C for 2.0 h to allow for the recovery of growth. Protein expression was induced by addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 1 mM. After induction for 2 h, the cells were harvested by centrifugation, and proteins were purified using Ni-NTA affinity chromatography as described above.

5.4.5 Affinity tag removal

To remove a polyhistidine tag, the proteins were subjected to subsequent TEV digestion. The protein buffer was changed to TEV digestion buffer (50 mM TRIS pH 8.0, 75 mM NaCl) using a BioRad 10 DG desalting column, and then dithiothreitol (DTT) and ethylenediaminetetraacetic acid (EDTA) were added to a final concentration of 1 and 0.5 mM, respectively. TEV protease was added to the protein, using 1:10 ratio of OD_{280}(TEV) to OD_{280}(protein). The solution was mixed, sterilized using a 0.22 μm filter, and incubated overnight at 34 °C. Following the digest, DTT and EDTA were removed with a desalting column. The obtained protein solution in 20 mM HEPES, 10 mM CaCl2, 100 mM NaCl (pH 7.5) was passed through a Ni-NTA column to remove the undigested protein and the cleaved His_6-tag.

5.5 Kinetic Assays

5.5.1 Kemp Assays

Kinetic measurements were done on a Thermo Labsystems Multiskan Spectrum plate reader monitoring absorbance of the product at 380 nm at 22 °C, in 200 μL reaction volumes, using 96-well plates. The reported results were obtained by averaging of at least three independent measurements. In a typical experiment 150 μL of substrate (0.12 -

0.96 mM in 20 mM HEPES (pH 7), 10 mM CaCl2, 100 mM NaCl) was added to 50 μL of buffered (0.8 μM) protein solution. 5-Nitrobenzisoxazole was used from a 0.1 M stock in acetonitrile (the acetonitrile percentage was 1.5 % in all reaction mixtures). The path length on the plate reader was determined empirically by measuring absorbance of the standard solution of the product on the plate reader and a UV-vis spectrophotometer. The empirically obtained conversion coefficient is then applied to the raw absorbance values obtained from the plate reader. Kinetic parameters were 30 readings (30 min) at 380 nm.

5.5.2 Esterase Hydrolytic Platereader Assay

To study the behavior of the protein, enzymes kinetics was done on a Thermo Labsystems Multiskan Spectrum platereader monitoring absorbance of the product at 405 nm, in 200 μL reaction volumes, using 96-well plates. Proteins were diluted to a final concentration of 2.0 μM and proteins were monitored in triplicate with various substrate concentrations of pNPP. Kinetic parameters were for 100 minutes, 300 readings every 20 seconds at 405 nm. Conducting a simple kinetic assay consist of 150 μL of varying substrate concentrations in 20 mM HEPES (pH 7.5), 10 mM $CaCl_2$, 100 mM NaCl, acetonitrile) added to 50 μL of buffered (~8.0 μM) protein solution for a final protein concentration of 2.0 μM. Substrate pNPP was used from a 10 mM stock in acetonitrile. The extinction coefficient of 15,800 $cm^{-1}M^{-1}$ was used to determine the product concentration at pH of 5.5 and above. At pH 4.5 and 5 the product extinction coefficients of 12,480 cm-1M-1 and 14,220 $cm^{-1}M^{-1}$, respectively, were used. Kinetic parameters for proteins (k_{cat} and K_M) were obtained by fitting the data to the Michaelis-Menten equation

$\{v_o = k_{cat}[E]_o[S]_o/(K_M+[S_o]$. k_{cat}/K_M values were calculated by obtaining the linear burst and slow phase rates and fitting them to $v_o = (k_{cat}/K_M)[E]_o[S]_o$ equation.

5.5.3 Esterase UV-Vis Assay

Kinetic assays were performed in a 1 cm path length quartz cuvette on the Agilent Cary 60 UV–vis spectrophotometer monitoring absorbance of the product at 405 nm. The extinction coefficient for p-nitrophenol (12,700 $M^{-1}cm^{-1}$) under these conditions was taken from the literature data. Proteins were diluted to a final concentration of 2.0 or 6.0 µM for hydrolysis of pNPP (R and S isomers) and pNPA, respectively. Substrate stock solutions of pNPP (1 mM) and pNPA (100 mM) were prepared in acetonitrile. The reaction was monitored in triplicate for 900 s (0.100 s averaging time) in 200 µL of buffer containing 20 mM HEPES (pH 7.5), 10 mM CaCl2, 100 mM NaCl. A 150 µL portion of substrate solution in buffer was added first to the cuvette, and subsequently, 50 µL of the protein solution was added. The final content of acetonitrile was 1.5% in reaction mixtures containing pNPA and 6% in the reaction mixtures containing pNPP. The obtained kinetic traces were analyzed using the method of Kezdy and Bender. Briefly, the kinetic traces were fit to

$$[p\text{-nitrophenolate}] = At + B(1 - e{-}bt) \quad (1)$$

where t is time. Plotting reciprocal values of A and b versus $1/[S]_o$ at various initial substrate concentrations allows for straightforward determination of k_{cat}, $K_{M(app)}$, k_2, and K_s.

5.5.4 pH Profiles

pH profiles were done with 2x buffers 50 mM citrate, MES, HEPES, TRIS, TAPS (0.5 M stock), 20 mM CaCl2 (1 M stock), 200 mM NaCl (3 M stock), and dd H2O added up to 20 mL at varying pH's 4.5-9: citrate (pH 4.5-5.5), MES (pH 6-6.5), HEPES (pH 7-7.5), TRIS (pH 8-8.5), TAPS (pH 9-9.5). Proteins were diluted in the same way as for kinetics to a final concentration of 0.2 μM with ddH$_2$O. Proteins were done in triplicate; blank ddH$_2$O in duplicate. Substrate was prepared 10 mL of 600 μM KEMP (9,800 μL H$_2$O, 60 μL of 100 mM KEMP, 140 μL of CH_3CN) to a final substrate concentration of 150 μM. Concentration of acetonitrile was 2% in the reaction mixture. The maximum k_{cat}/K_M values were obtained from fitting the pH dependence data to $k_{cat}/K_M = (k_{cat}/K_M)_{protonated} + (k_{cat}/K_M)_{deprotonated} \times 10^{-pKa}/(10^{-pH}+10^{-pKa})$, where pK_a is the apparent pK_a value of the active residue. Substrate was prepared 10 mL of 80 μM pNPP (9,500 μL H$_2$O, 80 μL of 10 mM KEMP, 420 μL of CH_3CN) to a final substrate concentration of 20 μM. Concentration of acetonitrile was 5% in the reaction mixture.

5.5.5 Calcium Dependence Assays

Ethylenediamintetraacetic acid (EDTA) was added to crude protein at a final concentration of 15 mM. Buffer exchange was completed twice into pH= 7 (20 mM HEPES, 100 mM NaCl). Different calcium concentrations used were 10 μM, 20, 50, 100, 200, 500, and 1 mM. EDTA at a final concentration of 5 μM was used as a control. Final concentrations were 0.2 μM proteins and 100 μM substrate (9800 μL buffer + 160 μL CH3CN + 40 μL of 100 mM Kemp). Reaction were completed for 20 min at 380 nm. Data for the dependence of the all CaM F92E 7th R proteins (0.20 μM) activity as a function of added metal concentration (0 - 1 mM of CaCl2). The data were fit to the Hill equation:

$$\frac{\frac{k_{cat}}{K_M}}{\left(\frac{k_{cat}}{K_M}\right)_{Max}} = \frac{\left(\frac{[M]_{tot}}{[M]_{50}}\right)^h}{(1 + \left(\frac{[M]_{tot}}{[M]_{50}}\right)^h}$$

where k_{cat}/K_M is enzymatic efficiency at the total metal concentration of $[M]_{tot}$, h is the Hill coefficient, M_{50} is the metal concentration that induces 50% enzymatic activity, and $(k_{cat}/K_M)_{max}$ is maximum enzymatic efficiency. To show that AlleyCatE is allosterically regulated in the presence of Ca^{2+} for its activity and could be completely inactivated in the absence of Ca^{2+}, an additional experiment was performed. Proteins were diluted to 2.0 μM and R-pNPP at the highest concentration of 50.0 μM in (20 mM HEPES, 100 mM NaCl, pH = 7.5). The sample was tested in the presence and absence of 10 mM Ca^{2+} and (EDTA) was added sequentially to a final 50 μM. Kinetic measurements were recorded using Multiskan spectrum monitoring absorbance of the product at 405 nm for 20 min, 200 μL reaction volumes on UV-Vis spectrophotometer.

5.6 Circular Dichroism Spectroscopy

Circular dichroism experiments were performed on a JASCO J-715 CD spectrometer in a step scan mode (4 s averaging time) averaging over three runs using a quartz cuvette with a 1 mm path length. Samples were prepared in a buffer containing 4 mM HEPES (pH 7.5), 2 mM CaCl2, and 30 mM NaCl with final protein concentration maintained at 25.0 μM. The sample absorbance never exceeded 1.5 at all wavelengths in order to obtain reliable MRE values. The ellipticity of proteins was monitored at 222 nm for chemical denaturation experiments in the presence of varying concentrations of guanidinium hydrochloride (0–6 M). The data collected for c-CaM and AlleyCatE were fitted to the equation below.

$$MRE = MRE_f + \frac{(MRE_u + y_u[D])e^{\frac{(-\Delta G - m[D])}{RT}}}{1 + e^{\frac{(-\Delta G - m[D])}{RT}}}$$

Here MRE is the observed mean residue ellipticity; MRE_f and MREu are mean residue ellipticities representing the folded and unfolded states for the C-terminal domain, respectively. ΔG is the free energy of unfolding; [D] is the concentration of the denaturant, and y_u is the slope for the unfolded state.

5.7 NMR Spectroscopy

The NMR samples of U-15N,13C -labeled protein were prepared at a concentration 1.1 mM in a buffer containing 5% D2O, 0.05% DSS, 0.01% NaN3. All experiments were conducted at 30 °C on a Varian 600 MHz spectrometer equipped with cryogenic probe. Chemical shift assignments were made using ^{15}N- and ^{13}C-HSQC, HNCO, HNCACB, CBCA(CO)HN, HBHA(CO)HN, 3D (H)CCHTOCSY. The assignments were further confirmed by 3D 15N- and 13C-edited NOESY spectra with the mixing time of 100 ms. NMR data were processed, analyzed, and visualized using NMRPipe and CARA. 1H chemical shifts were referenced to internal DSS, and ^{13}C and ^{15}N chemical shifts were referenced indirectly via gyromagnetic ratios. Backbone dihedral angle constraints were derived from chemical shifts using TALOS+.[109] 1H–1H upper distance constraints for structure calculations were obtained from the analysis of the 3D ^{13}C- and ^{15}Nedited NOESY spectra. Automated NOE assignment and structure calculation were done with CYANA 3.0[110] together with dihedral angle constraints and metal ion constraints, which were taken from the crystal structure of CaM (PDB code 1CLL).[73] The resulting assignments were verified and corrected by interactive spectral analysis. Subsequently, the structure calculation was performed iteratively with CYANA, with iterations used to verify and complete resonance assignments, refine NOESY peak lists, and optimize the distance calibration constants. The first set of refined structures was obtained using XPLOR-NIH.[111] The final set of structures was further refined by restrained molecular dynamics in explicit water[112] using the program CNS 1.2 with the OPLSX force field. Structural statistics and global quality factors were computed with PSVS 1.5.[113-116] Chemical shifts, NOESY peak lists, and raw time-domain data for AlleyCatE were deposited in the BioMagResBank (BMRB ID 19376); the coordinates were deposited in the Protein Data Bank (PDB code 4BYA).

5.8 Mass Spectrometry

MALDI-TOF spectrometry was done on a Bruker Autoflex III instrument. Prior to each experiment the instrument was calibrated with standard samples provided by themanufacturer. Dihydroxybenzoic acid (DHB) or sinapinic acid was used as a matrix. For acylated samples the substrate was added to the proteins, and the samples were immediately dried and analyzed.

5.9 Sequences

Proteins	Mutations	Sequence
c-CaM	none	SLMKDTDSEE EIREAFRVFD KDGNGYISAA ELRHVMTNLG EKLTDEEVDE MIREADIDGD GQVNYEEFVQ MMTAK*
AlleyCatE	M144H	SLMKDTDSEE EIREAFRVFD KDGNGYISAA ELRHVMTNLG EKLTDEEVDE MIREADIDGD GQVNYEEFVQ **H**MTAK*
AlleyCatE2	M144H, A128T	SLMKDTDSEE EIREAFRVFD KDGNGYISAA ELRHVMTNLG EKLTDEEVDE MIRE**T**DIDGD GQVNYEEFVQ HMTAK*

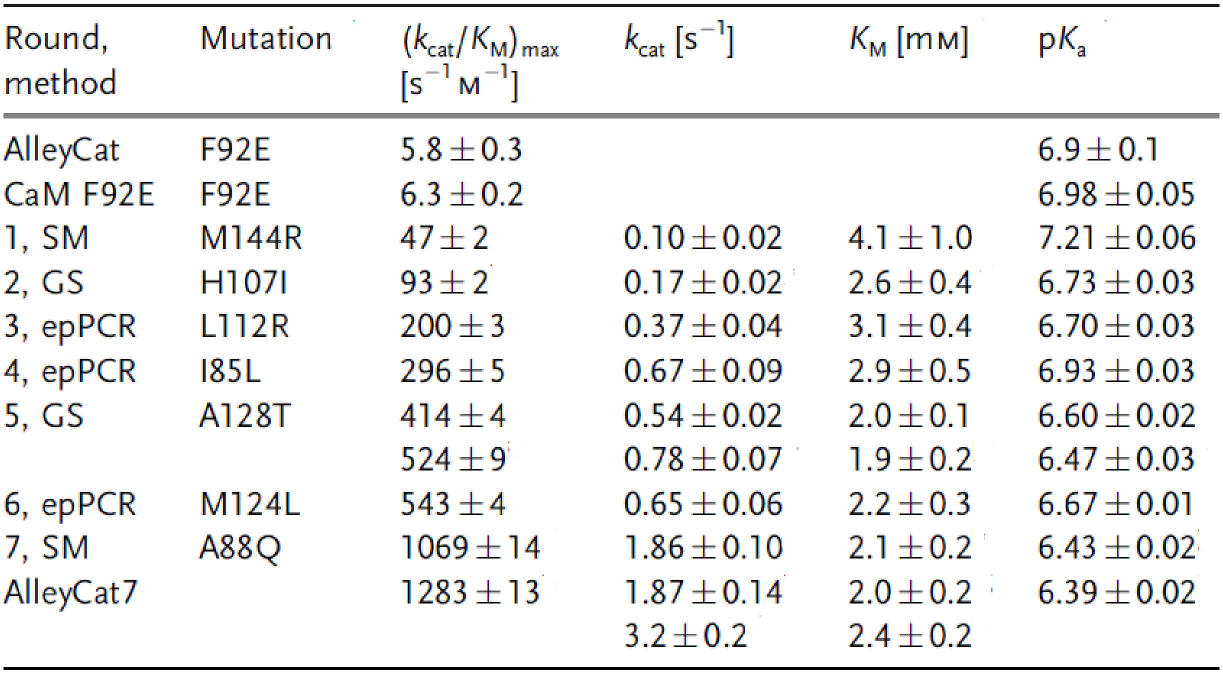
CaM F92E 7R

MSGSHHHHHH GSSGENLYFQ SLMADQLTEE QIAEFKEAFS LFDKDGDGTI
TTKELGTVMR SLGQNPTEAE LQDMINEVDA DGNGTIDFPE FLTMMARKMK
DTDSEEE**L**RE **Q**FRV**E**DKDGN GYISAAELR**I** VMTN**R**GEKLT DEEVDE**L**IRE
TDIDGDGQVN YEEFVQ**R**MTA K*

KE07 10/11G

MALAKRIDAALIMKDGRVVKGSNFENLRDSGDPVELGKFYSEIGIDELSFWDITASVEKRKT
MLELVEKVAEQIDIPITVGGGIHDFETASELILRGADKVEFNTAAVENPSLITQIAQTFGSQA
VVVYIAAKRVDGEFMVFTYSGTKNTGILLRDWVVEVEKRGAGEIVLGSIDRLGTKSGYDTE
MIRFVRPLTTLPIIAHRGAGKMEHFLEAFLAGADAAKADSVFHSREIDVRELKEYLKKHGV
NVRLEGLGSLEHHHHHH

Table 2.1: Catalytic efficiencies of F92E evolved variants[71]



Round, method	Mutation	$(k_{cat}/K_M)_{max}$ [s^{-1} M^{-1}]	k_{cat} [s^{-1}]	K_M [mM]	pK_a
AlleyCat	F92E	5.8 ± 0.3			6.9 ± 0.1
CaM F92E	F92E	6.3 ± 0.2			6.98 ± 0.05
1, SM	M144R	47 ± 2	0.10 ± 0.02	4.1 ± 1.0	7.21 ± 0.06
2, GS	H107I	93 ± 2	0.17 ± 0.02	2.6 ± 0.4	6.73 ± 0.03
3, epPCR	L112R	200 ± 3	0.37 ± 0.04	3.1 ± 0.4	6.70 ± 0.03
4, epPCR	I85L	296 ± 5	0.67 ± 0.09	2.9 ± 0.5	6.93 ± 0.03
5, GS	A128T	414 ± 4	0.54 ± 0.02	2.0 ± 0.1	6.60 ± 0.02
		524 ± 9	0.78 ± 0.07	1.9 ± 0.2	6.47 ± 0.03
6, epPCR	M124L	543 ± 4	0.65 ± 0.06	2.2 ± 0.3	6.67 ± 0.01
7, SM	A88Q	1069 ± 14	1.86 ± 0.10	2.1 ± 0.2	6.43 ± 0.02
AlleyCat7		1283 ± 13	1.87 ± 0.14	2.0 ± 0.2	6.39 ± 0.02
			3.2 ± 0.2	2.4 ± 0.2	

References

1. Richard, J. P., Enzymatic Rate Enhancements: A Review and Perspective(). *Biochemistry* **2013,** *52* (12), 10.1021/bi3017119.

2. Herschlag, D.; Natarajan, A., Fundamental Challenges in Mechanistic Enzymology: Progress toward Understanding the Rate Enhancements of Enzymes. *Biochemistry* **2013,** *52* (12), 10.1021/bi4000113.

3. Agarwal, P. K., Enzymes: An integrated view of structure, dynamics and function. *Microbial Cell Factories* **2006,** *5* (1), 2.

4. Cleland, W. W., What limits the rate of an enzyme-catalyzed reaction. *Accounts of Chemical Research* **1975,** *8* (5), 145-151.

5. Robinson, R., What Governs Enzyme Activity? For One Enzyme, Charge Contributes Only Weakly. *PLOS Biology* **2006,** *4* (4), e133.

6. Gluza, K.; Kafarski, P., *Transition State Analogues of Enzymatic Reaction as Potential Drugs*. 2013.

7. Kastle, J. H., on the vital activity of enzymes. *Science* **1901,** *13* (333), 765-771.

8. Levene, P. A., The chemical nature of enzymes. *Journal of the American Chemical Society* **1901,** *23* (7), 505-508.

9. Cohen, S. G.; Khedouri, E., Requirements for Stereospecificity in Hydrolysis by α-Chymotrypsin. IV. The Hydroxyl Substituent. Absolute Configurations1. *Journal of the American Chemical Society* **1961,** *83* (20), 4228-4233.

10. H E Sutton, a.; Wagner, R. P., Mutation and Enzyme Function in Humans. *Annual Review of Genetics* **1975,** *9* (1), 187-212.

11. Lu, Y.; Berry, S. M.; Pfister, T. D., Engineering Novel Metalloproteins: Design of Metal-Binding Sites into Native Protein Scaffolds. *Chemical Reviews* **2001,** *101* (10), 3047-3080.

12. Lu, Y.; Yeung, N.; Sieracki, N.; Marshall, N. M., Design of functional metalloproteins. *Nature* **2009,** *460* (7257), 855-862.

13. Palm-Espling, M. E.; Niemiec, M. S.; Wittung-Stafshede, P., Role of metal in folding and stability of copper proteins in vitro. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **2012,** *1823* (9), 1594-1603.

14. Peacock, A. F. A., Incorporating metals into de novo proteins. *Current Opinion in Chemical Biology* **2013,** *17* (6), 934-939.

15. Höhne, M.; Bornscheuer, U. T., Protein Engineering from “Scratch” Is Maturing. *Angewandte Chemie International Edition* **2014,** *53* (5), 1200-1202.

16. Chica, R. A.; Doucet, N.; Pelletier, J. N., Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design. *Curr. Opin. Biotechnol.* **2005,** *16*, 378-384.

17. Kaplan, J.; DeGrado, W. F., De novo design of catalytic proteins. *Proc. Natl Acad. Sci. USA* **2004,** *101*, 11566-11570.

18. Park, H.-S.; Nam, S.-H.; Lee, J. K.; Yoon, C. N.; Mannervik, B.; Benkovic, S. J.; Kim, H.-S., Design and Evolution of New Catalytic Activity with an Existing Protein Scaffold. *Science* **2006,** *311* (5760), 535-538.
19. Head-Gordon, T.; Brown, S., Minimalist models for protein folding and design. *Current Opinion in Structural Biology* **2003,** *13* (2), 160-167.
20. Heinisch, T.; Ward, T. R., Design strategies for the creation of artificial metalloenzymes. *Current Opinion in Chemical Biology* **2010,** *14* (2), 184-199.
21. Marks, D. S.; Hopf, T. A.; Sander, C., Protein structure prediction from sequence variation. *Nature biotechnology* **2012,** *30* (11), 1072-1080.
22. Bornscheuer, U. T.; Pohl, M., Improved biocatalysts by directed evolution and rational protein design. *Current Opinion in Chemical Biology* **2001,** *5* (2), 137-143.
23. Bolon, D. N.; Mayo, S. L., Enzyme-like proteins by computational design. *Proc. Natl Acad. Sci. USA* **2001,** *98*, 14274-14279.
24. Radzicka, A.; Wolfenden, R., A proficient enzyme. *Science* **1995,** *267*, 90-93.
25. Thorn, S. N.; Daniels, R. G.; Auditor, M. T.; Hilvert, D., Large rate accelerations in antibody catalysis by strategic use of haptenic charge. *Nature* **1995,** *373*, 228-230.
26. Zanghellini, A., New algorithms and an in silico benchmark for computational enzyme design. *Protein Sci.* **2006,** *15*, 2785-2794.
27. Bloom, J. D.; Labthavikul, S. T.; Otey, C. R.; Arnold, F. H., Protein stability promotes evolvability. *Proceedings of the National Academy of Sciences* **2006,** *103* (15), 5869-5874.
28. Custódio, F. L.; Barbosa, H. J. C.; Dardenne, L. E., A multiple minima genetic algorithm for protein structure prediction. *Applied Soft Computing* **2014,** *15*, 88-99.
29. Radivojac, P.; Clark, W. T.; Oron, T. R.; Schnoes, A. M.; Wittkop, T.; Sokolov, A.; Graim, K.; Funk, C.; Verspoor, K.; Ben-Hur, A.; Pandey, G.; Yunes, J. M.; Talwalkar, A. S.; Repo, S.; Souza, M. L.; Piovesan, D.; Casadio, R.; Wang, Z.; Cheng, J.; Fang, H.; Gough, J.; Koskinen, P.; Toronen, P.; Nokso-Koivisto, J.; Holm, L.; Cozzetto, D.; Buchan, D. W. A.; Bryson, K.; Jones, D. T.; Limaye, B.; Inamdar, H.; Datta, A.; Manjari, S. K.; Joshi, R.; Chitale, M.; Kihara, D.; Lisewski, A. M.; Erdin, S.; Venner, E.; Lichtarge, O.; Rentzsch, R.; Yang, H.; Romero, A. E.; Bhat, P.; Paccanaro, A.; Hamp, T.; Kaszner, R.; Seemayer, S.; Vicedo, E.; Schaefer, C.; Achten, D.; Auer, F.; Boehm, A.; Braun, T.; Hecht, M.; Heron, M.; Honigschmid, P.; Hopf, T. A.; Kaufmann, S.; Kiening, M.; Krompass, D.; Landerer, C.; Mahlich, Y.; Roos, M.; Bjorne, J.; Salakoski, T.; Wong, A.; Shatkay, H.; Gatzmann, F.; Sommer, I.; Wass, M. N.; Sternberg, M. J. E.; Skunca, N.; Supek, F.; Bosnjak, M.; Panov, P.; Dzeroski, S.; Smuc, T.; Kourmpetis, Y. A. I.; van Dijk, A. D. J.; Braak, C. J. F. t.; Zhou, Y.; Gong, Q.; Dong, X.; Tian, W.; Falda, M.; Fontana, P.; Lavezzo, E.; Di Camillo, B.; Toppo, S.; Lan, L.; Djuric, N.; Guo, Y.; Vucetic, S.; Bairoch, A.; Linial, M.; Babbitt, P. C.; Brenner, S. E.; Orengo, C.; Rost, B.; Mooney, S. D.; Friedberg, I., A large-scale evaluation of computational protein function prediction. *Nat Meth* **2013,** *10* (3), 221-227.
30. Meiler, J.; Baker, D., ROSETTALIGAND: protein-small molecule docking with full side-chain flexibility. *Proteins* **2006,** *65*, 538-548.
31. Leaver-Fay, A.; Tyka, M.; Lewis, S. M.; Lange, O. F.; Thompson, J.; Jacak, R.; Kaufman, K. W.; Renfrew, P. D.; Smith, C. A.; Sheffler, W.; Davis, I. W.; Cooper, S.; Treuille, A.; Mandell, D. J.; Richter, F.; Ban, Y.-E. A.; Fleishman, S. J.; Corn, J. E.; Kim, D.

E.; Lyskov, S.; Berrondo, M.; Mentzer, S.; Popović, Z.; Havranek, J. J.; Karanicolas, J.; Das, R.; Meiler, J.; Kortemme, T.; Gray, J. J.; Kuhlman, B.; Baker, D.; Bradley, P., Chapter nineteen - Rosetta3: An Object-Oriented Software Suite for the Simulation and Design of Macromolecules. In *Methods in Enzymology*, Michael, L. J.; Ludwig, B., Eds. Academic Press: 2011; Vol. Volume 487, pp 545-574.
32. Koradi, R.; Billeter, M.; Wüthrich, K., MOLMOL: A program for display and analysis of macromolecular structures. *Journal of Molecular Graphics* **1996,** *14* (1), 51-55.
33. Kolinski, A.; Skolnick, J., Monte carlo simulations of protein folding. I. Lattice model and interaction scheme. *Proteins: Structure, Function, and Bioinformatics* **1994,** *18* (4), 338-352.
34. Currin, A.; Swainston, N.; Day, P. J.; Kell, D. B., Synthetic biology for the directed evolution of protein biocatalysts: navigating sequence space intelligently. *Chemical Society Reviews* **2015,** *44* (5), 1172-1239.
35. Brissos, V.; Gonçalves, N.; Melo, E. P.; Martins, L. O., Improving Kinetic or Thermodynamic Stability of an Azoreductase by Directed Evolution. *PLoS ONE* **2014,** *9* (1), e87209.
36. Eijsink, V. G. H.; Gåseidnes, S.; Borchert, T. V.; van den Burg, B., Directed evolution of enzyme stability. *Biomolecular Engineering* **2005,** *22* (1–3), 21-30.
37. Reetz, M. T., Introduction to Directed Evolution. In *Directed Evolution of Selective Enzymes*, Wiley-VCH Verlag GmbH & Co. KGaA: 2016; pp 1-25.
38. McLachlan, M. J.; Sullivan, R. P.; Zhao, H., Directed Enzyme Evolution and High-Throughput Screening. In *Biocatalysis for the Pharmaceutical Industry*, John Wiley & Sons, Ltd: 2009; pp 45-64.
39. Kunkel, T. A.; Roberts, J. D.; Zakour, R. A., Rapid and efficient site-specific mutagenesis without phenotypic selection. *Methods Enzymol.* **1987,** *154*, 367-382.
40. Packer, M. S.; Liu, D. R., Methods for the directed evolution of proteins. *Nat Rev Genet* **2015,** *16* (7), 379-394.
41. Li, Z.; Chen, X.; Guo, S.; Zhang, H.; Dong, H.; Wu, G.; Hong, A.; Gu, J., Engineering of Harobin for enhanced fibrinolytic activity obtained by random and site-directed mutagenesis. *Protein Expression and Purification* **2017,** *129*, 162-172.
42. Head, S. R.; Komori, H. K.; LaMere, S. A.; Whisenant, T.; Van Nieuwerburgh, F.; Salomon, D. R.; Ordoukhanian, P., Library construction for next-generation sequencing: Overviews and challenges. *BioTechniques* **2014,** *56* (2), 61-passim.
43. Tyagi, R.; Lai, R.; Duggleby, R. G., A new approach to 'megaprimer' polymerase chain reaction mutagenesis without an intermediate gel purification step. *BMC Biotechnology* **2004,** *4*, 2-2.
44. Israel, D. I., A PCR-based method for high stringency screening of DNA libraries. *Nucleic Acids Research* **1993,** *21* (11), 2627-2631.
45. Acevedo-Rocha, C. G.; Reetz, M. T.; Nov, Y., Economical analysis of saturation mutagenesis experiments. *Scientific Reports* **2015,** *5*, 10654.
46. Mena, M. A.; Daugherty, P. S., Automated design of degenerate codon libraries. *Protein Engineering Design and Selection* **2005,** *18* (12), 559-561.
47. Studier, F. W., Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **2005,** *41*, 207-234.

48. Cabrita, L. D.; Dai, W.; Bottomley, S. P., A family of E. coli expression vectors for laboratory scale and high throughput soluble protein production. *BMC Biotechnol.* **2006,** *6*, 12.

49. Kapust, R. B.; Waugh, D. S., Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci.* **1999,** *8*, 1668-1674.

50. Rosano, G. L.; Ceccarelli, E. A., Recombinant protein expression in Escherichia coli: advances and challenges. *Front Microbiol* **2014,** *5*, 172.

51. Rong Fu, W.; Kushner, S. R., Construction of versatile low-copy-number vectors for cloning, sequencing and gene expression in Escherichia coli. *Gene* **1991,** *100*, 195-199.

52. Tabor, S., Expression Using the T7 RNA Polymerase/Promoter System. In *Current Protocols in Molecular Biology*, John Wiley & Sons, Inc.: 2001.

53. Sivashanmugam, A.; Murray, V.; Cui, C.; Zhang, Y.; Wang, J.; Li, Q., Practical protocols for production of very high yields of recombinant proteins using Escherichia coli. *Protein Science : A Publication of the Protein Society* **2009,** *18* (5), 936-948.

54. Butt, T. R.; Edavettal, S. C.; Hall, J. P.; Mattern, M. R., SUMO fusion technology for difficult-to-express proteins. *Protein Expression and Purification* **2005,** *43* (1), 1-9.

55. Routzahn, K. M.; Waugh, D. S., Differential effects of supplementary affinity tags on the solubility of MBP fusion proteins. *J. Struct. Funct. Genomics* **2002,** *2*, 83-92.

56. Hristozova, N.; Tompa, P.; Kovacs, D., A Novel Method for Assessing the Chaperone Activity of Proteins. *PLOS ONE* **2016,** *11* (8), e0161970.

57. Nallamsetty, S.; Waugh, D. S., Solubility-enhancing proteins MBP and NusA play a passive role in the folding of their fusion partners. *Protein Expr. Purif.* **2006,** *45*, 175-182.

58. Scopes, R. K., Protein Purification. 1994.

59. Structural Genomics, C.; Architecture et Fonction des Macromolécules, B.; Berkeley Structural Genomics, C.; China Structural Genomics, C.; Integrated Center for, S.; Function, I.; Israel Structural Proteomics, C.; Joint Center for Structural, G.; Midwest Center for Structural, G.; New York Structural Genomi, X. R. C. f. S. G.; Northeast Structural Genomics, C.; Oxford Protein Production, F.; Protein Sample Production Facility, M. D. C. f. M. M.; Initiative, R. S. G. P.; Complexes, S., Protein production and purification. *Nature methods* **2008,** *5* (2), 135-146.

60. Tropea, J. E.; Cherry, S.; Waugh, D. S., Expression and purification of soluble His6-tagged TEV protease. *Methods. Mol. Biol.*

61. Waugh, D. S., Making the most of affinity tags. *Trends Biotechnol.* **2005,** *23*, 316-320.

62. Zhu, K.; Zhou, X.; Yan, Y.; Mo, H.; Xie, Y.; Cheng, B.; Fan, J., Cleavage of fusion proteins on the affinity resins using the TEV protease variant. *Protein Expression and Purification* **2017,** *131*, 27-33.

63. Zastrow, M. L.; Pecoraro, V. L., Designing functional metalloproteins: From structural to catalytic metal sites. *Coordination Chemistry Reviews* **2013,** *257* (17–18), 2565-2588.

64. Lu, Y., Design and engineering of metalloproteins containing unnatural amino acids or non-native metal-containing cofactors. *Current Opinion in Chemical Biology* **2005,** *9* (2), 118-126.

65. DeGrado, W.; Wasserman, Z.; Lear, J., Protein design, a minimalist approach. *Science* **1989,** *243* (4891), 622-628.

66. Dantas, G.; Kuhlman, B.; Callender, D.; Wong, M.; Baker, D., A large scale test of computational protein design: folding and stability of nine completely redesigned globular proteins. *J. Mol. Biol.* **2003,** *332*, 449-460.

67. Kiss, G.; Çelebi-Ölçüm, N.; Moretti, R.; Baker, D.; Houk, K. N., Computational Enzyme Design. *Angewandte Chemie International Edition* **2013,** *52* (22), 5700-5725.

68. Khersonsky, O.; Kiss, G.; Röthlisberger, D.; Dym, O.; Albeck, S.; Houk, K. N.; Baker, D.; Tawfik, D. S., Bridging the gaps in design methodologies by evolutionary optimization of the stability and proficiency of designed Kemp eliminase KE59. *Proceedings of the National Academy of Sciences* **2012,** *109* (26), 10358-10363.

69. Feng, X.; Sanchis, J.; Reetz, M. T.; Rabitz, H., Enhancing the Efficiency of Directed Evolution in Focused Enzyme Libraries by the Adaptive Substituent Reordering Algorithm. *Chemistry – A European Journal* **2012,** *18* (18), 5646-5654.

70. Korendovych, I. V.; Kulp, D. W.; Wu, Y.; Cheng, H.; Roder, H.; DeGrado, W. F., Design of a switchable eliminase. *Proc Natl Acad Sci U S A* **2011,** *108* (17), 6823-6827.

71. Moroz, O. V.; Moroz, Y. S.; Wu, Y.; Olsen, A. B.; Cheng, H.; Mack, K. L.; McLaughlin, J. M.; Raymond, E. A.; Zhezherya, K.; Roder, H.; Korendovych, I. V., A Single Mutation in a Regulatory Protein Produces Evolvable Allosterically Regulated Catalyst of Nonnatural Reaction. *Angewandte Chemie International Edition* **2013,** *52* (24), 6246-6249.

72. Moroz, Y. S.; Dunston, T. T.; Makhlynets, O. V.; Moroz, O. V.; Wu, Y.; Yoon, J. H.; Olsen, A. B.; McLaughlin, J. M.; Mack, K. L.; Gosavi, P. M.; van Nuland, N. A. J.; Korendovych, I. V., New Tricks for Old Proteins: Single Mutations in a Nonenzymatic Protein Give Rise to Various Enzymatic Activities. *Journal of the American Chemical Society* **2015,** *137* (47), 14905-14911.

73. Chattopadhyaya, R.; Meador, W. E.; Means, A. R.; Quiocho, F. A., Calmodulin structure refined at 1.7 Å resolution. *Journal of Molecular Biology* **1992,** *228* (4), 1177-1192.

74. Blomberg, R.; Kries, H.; Pinkas, D. M.; Mittl, P. R. E.; Grutter, M. G.; Privett, H. K.; Mayo, S. L.; Hilvert, D., Precision is essential for efficient catalysis in an evolved Kemp eliminase. *Nature* **2013,** *503* (7476), 418-421.

75. Kemp, D. S.; Casey, M. L., Physical organic chemistry of benzisoxazoles II. Linearity of the bronsted free energy relationship for the base-catalyzed decomposition of benzisoxazoles. *J. Am. Chem. Soc.* **1973,** *95*, 6670-6680.

76. Rothlisberger, D.; Khersonsky, O.; Wollacott, A. M.; Jiang, L.; DeChancie, J.; Betker, J.; Gallaher, J. L.; Althoff, E. A.; Zanghellini, A.; Dym, O.; Albeck, S.; Houk, K. N.; Tawfik, D. S.; Baker, D., Kemp elimination catalysts by computational enzyme design. *Nature* **2008,** *453* (7192), 190-195.

77. Johnson, K. A.; Goody, R. S., The Original Michaelis Constant: Translation of the 1913 Michaelis–Menten Paper. *Biochemistry* **2011,** *50* (39), 8264-8269.

78. Fujii, R.; Kitaoka, M.; Hayashi, K., One-step random mutagenesis by error-prone rolling circle amplification. *Nucleic Acids Research* **2004,** *32* (19), e145.

79. Fife, T. H.; McMahon, D. M., Acid- and water-catalyzed hydrolysis of p-nitrophenyl esters. *Journal of the American Chemical Society* **1969,** *91* (26), 7481-7485.

80. Engström, K.; Nyhlén, J.; Sandström, A. G.; Bäckvall, J.-E., Directed Evolution of an Enantioselective Lipase with Broad Substrate Scope for Hydrolysis of α-Substituted Esters. *Journal of the American Chemical Society* **2010,** *132* (20), 7038-7042.
81. Khare, S. D.; Kipnis, Y.; Greisen, P., Jr.; Takeuchi, R.; Ashani, Y.; Goldsmith, M.; Song, Y.; Gallaher, J. L.; Silman, I.; Leader, H.; Sussman, J. L.; Stoddard, B. L.; Tawfik, D. S.; Baker, D., Computational redesign of a mononuclear zinc metalloenzyme for organophosphate hydrolysis. *Nat Chem Biol* **2012,** *8* (3), 294-300.
82. Linder, M., COMPUTATIONAL ENZYME DESIGN: ADVANCES, HURDLES AND POSSIBLE WAYS FORWARD. *Computational and Structural Biotechnology Journal* **2012,** *2* (3), 1-8.
83. Raymond, E. A.; Mack, K. L.; Yoon, J. H.; Moroz, O. V.; Moroz, Y. S.; Korendovych, I. V., Design of an allosterically regulated retroaldolase. *Protein Science : A Publication of the Protein Society* **2015,** *24* (4), 561-570.
84. Richter, F.; Blomberg, R.; Khare, S. D.; Kiss, G.; Kuzin, A. P.; Smith, A. J. T.; Gallaher, J.; Pianowski, Z.; Helgeson, R. C.; Grjasnow, A.; Xiao, R.; Seetharaman, J.; Su, M.; Vorobiev, S.; Lew, S.; Forouhar, F.; Kornhaber, G. J.; Hunt, J. F.; Montelione, G. T.; Tong, L.; Houk, K. N.; Hilvert, D.; Baker, D., Computational Design of Catalytic Dyads and Oxyanion Holes for Ester Hydrolysis. *Journal of the American Chemical Society* **2012,** *134* (39), 16197-16206.
85. Behrens, G. A.; Hummel, A.; Padhi, S. K.; Schätzle, S.; Bornscheuer, U. T., Discovery and Protein Engineering of Biocatalysts for Organic Synthesis. *Advanced Synthesis & Catalysis* **2011,** *353* (13), 2191-2215.
86. Alvira, P.; Tomás-Pejó, E.; Ballesteros, M.; Negro, M. J., Pretreatment technologies for an efficient bioethanol production process based on enzymatic hydrolysis: A review. *Bioresource Technology* **2010,** *101* (13), 4851-4861.
87. Trott, O.; Olson, A. J., AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry* **2010,** *31* (2), 455-461.
88. Aharoni, A.; Thieme, K.; Chiu, C. P. C.; Buchini, S.; Lairson, L. L.; Chen, H.; Strynadka, N. C. J.; Wakarchuk, W. W.; Withers, S. G., High-throughput screening methodology for the directed evolution of glycosyltransferases. *Nat Meth* **2006,** *3* (8), 609-614.
89. Schmidt, M.; Hasenpusch, D.; Kähler, M.; Kirchner, U.; Wiggenhorn, K.; Langel, W.; Bornscheuer, U. T., Directed Evolution of an Esterase from Pseudomonas fluorescens Yields a Mutant with Excellent Enantioselectivity and Activity for the Kinetic Resolution of a Chiral Building Block. *ChemBioChem* **2006,** *7* (5), 805-809.
90. Bender, M. L.; Kezdy, F. J.; Wedler, F. C., alpha-Chymotrypsin: Enzyme concentration and kinetics. *Journal of Chemical Education* **1967,** *44* (2), 84.
91. Tao, H.; Cornish, V. W., Milestones in directed enzyme evolution. *Current Opinion in Chemical Biology* **2002,** *6* (6), 858-864.
92. Banfield, M. J.; Lott, J. S.; Arcus, V. L.; McCarthy, A. A.; Baker, E. N., Structure of HisF, a histidine biosynthetic protein from Pyrobaculum aerophilum. *Acta crystallographica. Section D, Biological crystallography* **2001,** *57* (Pt 11), 1518-25.

93. Taglieber, A.; Höbenreich, H.; Carballeira, J. D.; Mondière, R. J. G.; Reetz, M. T., Alternate-Site Enzyme Promiscuity. *Angewandte Chemie International Edition* **2007,** *46* (45), 8597-8600.

94. Zastrow, M. L.; Pecoraro, V. L., Influence of Active Site Location on Catalytic Activity in de Novo-Designed Zinc Metalloenzymes. *Journal of the American Chemical Society* **2013,** *135* (15), 5895-5903.

95. Wierenga, R. K., The TIM-barrel fold: a versatile framework for efficient enzymes. *FEBS Letters* **2001,** *492* (3), 193-198.

96. Schmidt, D. M. Z.; Mundorff, E. C.; Dojka, M.; Bermudez, E.; Ness, J. E.; Govindarajan, S.; Babbitt, P. C.; Minshull, J.; Gerlt, J. A., Evolutionary Potential of (β/α)8-Barrels: Functional Promiscuity Produced by Single Substitutions in the Enolase Superfamily. *Biochemistry* **2003,** *42* (28), 8387-8393.

97. Khersonsky, O.; Röthlisberger, D.; Dym, O.; Albeck, S.; Jackson, C. J.; Baker, D.; Tawfik, D. S., Evolutionary Optimization of Computationally Designed Enzymes: Kemp Eliminases of the KE07 Series. *Journal of Molecular Biology* **2010,** *396* (4), 1025-1042.

98. Nagano, N.; Orengo, C. A.; Thornton, J. M., One Fold with Many Functions: The Evolutionary Relationships between TIM Barrel Families Based on their Sequences, Structures and Functions. *Journal of Molecular Biology* **2002,** *321* (5), 741-765.

99. Barik, S., Megaprimer PCR. In *PCR Cloning Protocols*, Chen, B.-Y.; Janes, H. W., Eds. Humana Press: Totowa, NJ, 2002; pp 189-196.

100. Pandya, C.; Farelli, J. D.; Dunaway-Mariano, D.; Allen, K. N., Enzyme Promiscuity: Engine of Evolutionary Innovation. *Journal of Biological Chemistry* **2014,** *289* (44), 30229-30236.

101. Hult, K.; Berglund, P., Enzyme promiscuity: mechanism and applications. *Trends in Biotechnology* **2007,** *25* (5), 231-238.

102. Leščić Ašler, I.; Ivić, N.; Kovačić, F.; Schell, S.; Knorr, J.; Krauss, U.; Wilhelm, S.; Kojić-Prodić, B.; Jaeger, K.-E., Probing Enzyme Promiscuity of SGNH Hydrolases. *ChemBioChem* **2010,** *11* (15), 2158-2167.

103. Lang, D.; Thoma, R.; Henn-Sax, M.; Sterner, R.; Wilmanns, M., Structural Evidence for Evolution of the β/α Barrel Scaffold by Gene Duplication and Fusion. *Science* **2000,** *289* (5484), 1546-1550.

104. Brooks, B. R.; Brooks, C. L.; MacKerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M., CHARMM: The Biomolecular Simulation Program. *Journal of computational chemistry* **2009,** *30* (10), 1545-1614.

105. Morris, G. M.; Huey, R.; Olson, A. J., Using AutoDock for Ligand-Receptor Docking. In *Current Protocols in Bioinformatics*, John Wiley & Sons, Inc.: 2002.

106. Hu, Y.; Houk, K. N.; Kikuchi, K.; Hotta, K.; Hilvert, D., Nonspecific Medium Effects versus Specific Group Positioning in the Antibody and Albumin Catalysis of the Base-Promoted Ring-Opening Reactions of Benzisoxazoles. *Journal of the American Chemical Society* **2004,** *126* (26), 8197-8205.

107. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E., UCSF Chimera—A visualization system for exploratory research and analysis. *Journal of Computational Chemistry* **2004,** *25* (13), 1605-1612.

108. Eschenfeldt, W. H.; Stols, L.; Millard, C. S.; Joachimiak, A.; Donnelly, M. I., A Family of LIC Vectors for High-Throughput Cloning and Purification of Proteins. *Methods in molecular biology (Clifton, N.J.)* **2009,** *498*, 105-115.

109. Shen, Y.; Delaglio, F.; Cornilescu, G.; Bax, A., TALOS+: A hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *Journal of biomolecular NMR* **2009,** *44* (4), 213-223.

110. Güntert, P., Automated NMR Structure Calculation With CYANA. In *Protein NMR Techniques*, Downing, A. K., Ed. Humana Press: Totowa, NJ, 2004; pp 353-378.

111. Schwieters, C. D.; Kuszewski, J. J.; Tjandra, N.; Marius Clore, G., The Xplor-NIH NMR molecular structure determination package. *Journal of Magnetic Resonance* **2003,** *160* (1), 65-73.

112. Linge, J. P.; Williams, M. A.; Spronk, C. A. E. M.; Bonvin, A. M. J. J.; Nilges, M., Refinement of protein structures in explicit solvent. *Proteins: Structure, Function, and Bioinformatics* **2003,** *50* (3), 496-506.

113. Luthy, R.; Bowie, J. U.; Eisenberg, D., Assessment of protein models with three-dimensional profiles. *Nature* **1992,** *356* (6364), 83-85.

114. Lovell, S. C.; Davis, I. W.; Arendall, W. B.; de Bakker, P. I. W.; Word, J. M.; Prisant, M. G.; Richardson, J. S.; Richardson, D. C., Structure validation by Cα geometry: ϕ,ψ and Cβ deviation. *Proteins: Structure, Function, and Bioinformatics* **2003,** *50* (3), 437-450.

115. Sippl, M. J., Recognition of errors in three-dimensional structures of proteins. *Proteins: Structure, Function, and Bioinformatics* **1993,** *17* (4), 355-362.

116. Pulavarti, S. V.; Eletsky, A.; Lee, H.-W.; Acton, T. B.; Xiao, R.; Everett, J. K.; Prestegard, J. H.; Montelione, G. T.; Szyperski, T., Solution NMR structure of CD1104B from pathogenic Clostridium difficile reveals a distinct α-helical architecture and provides first structural representative of protein domain family PF14203. *Journal of structural and functional genomics* **2013,** *14* (4), 10.1007/s10969-013-9164-8.

Tiffany Dunston

1820 Alabama Ave SE Washington, D.C 20020
ttdunsto@syr.edu, (301) 442-9295, www.linkedin.com/in/tiffany-dunston

EDUCATION

Ph.D. in Chemistry, **Syracuse University**, Syracuse NY **Expected 2017**

B.S. in Biochemistry, **Syracuse University**, Syracuse NY **2012**

RESEARCH EXPERIENCE

Graduate Fellow, Syracuse University **2012 - Present**

- Utilized a minimalistic protein design approach to create catalysts for chemical reactions such (Kemp elimination and ester hydrolysis)
- Recombinantly express DNA into E.*coli*
- Optimize protein expression and purification protocols

Intern, National Nanotechnology Infrastructure Network Program, Howard University, Washington DC **2011**

- Prepared dextran solutions of various molecular weights
- Measured the total carbon concentration of dextran permeate solutions using Total Organic Carbon (TOC) Analyzer
- Identified the membrane with the right molecular weight cutoff (MWCO) for bioseparation
- Imaged alumina collagen modified membranes using Scanning Electron Microscopy http://s3.amazonaws.com/nnin/pbbstream_1.mp4

RESEARCH SKILLS

- Polymerase Chain Reaction
- E.*coli* Transformation
- DNA electrophoresis
- Spectroscopic methods: UV-Vis spectroscopy, IR spectroscopy, and Proton NMR spectroscopy
- TLC chromatography

FELLOWSHIPS & AWARDS

- Syracuse University Arts & Sciences Stipend **2016**
- The Ellithorp Stafford Fellowship **2016**

- 14th Upstate New York NMR Conference & Structural Biology Symposium, Poster prize winner **2015**
- NSF Integrative Graduate Education and Research Traineeship (IGERT) Fellowship **2013**
- Graduate Assistance in areas of National Need (GAAN) Fellowship, Syracuse University **2012**

PEER REVIEWED PUBLICATION

- Moroz, Y. S.; **Dunston, T. T**.; Makhlynets, O. V.; Moroz, O. V.; et al. New Tricks for Old Proteins: Single Mutations in a Nonenzymatic Protein Give Rise to Various Enzymatic Activities. J. Am. Chem. Soc. **2015**, 137, 14905-14911. (**Featured in *JACS Spotlights***)

PRESENTATIONS

- 41st Northeast Regional Meeting, Binghamton, NY **2016**
  "Directed Evolution for the design of new catalysts" (talk)
- 252nd American Chemical Society National Meeting, Philadelphia, PA **2016**
  "New Tricks for old Proteins" (poster)
- IGERT Retreat Conference, Syracuse University, Syracuse, NY **2015**
  "*De novo* design of Catalytic Materials" (talk)
- 14th Upstate New York NMR Conference and Structural Biology Symposium "New tricks for Old Proteins" (poster) **2015**
- 33rd Annual Graduate Student Symposium, University at Buffalo, Buffalo,NY "Design of an Allosterically Regulated Esterase" (poster) **2015**
- IGERT Retreat Conference, Syracuse University **2014**
  "New Tricks for Old Proteins" (poster)

COMPUTER SKILLS

- Proficient at Microsoft Office
- Advanced at PYMOL, CHEMDRAW, 4PEAKS, and KALEIDAGRAPH software

LEADERSHIP & COMMUNICATION EXPERIENCE

Teaching Assistant of Organic Chemistry Laboratory (~60 students weekly) **2015,2016**

- Keynote speaker for School Choice in America Legislator Summit **2016**
- Prospective Incoming Graduate Student Visitation **2016**
- Women in Science and Engineering Panel **2016**
- IGERT Retreat **2015**
- The National Gem Consortium: GEM Lab **2015**
- McNair Scholar Graduate Panel **2014**